PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A61K 31/16, 31/165, 31/415, 31/425, 38/05, C12N 9/48, 9/64, C12Q 1/37

A1

(11) International Publication Number: WO 97/16177

(43) International Publication Date: 9 May 1997 (09.05.97)

(21) International Application Number: PCT/US96/17512

(22) International Filing Date: 30 October 1996 (30.10.96)

(30) Priority Data:

60/008,108	30 October 1995 (30.10.95)	US
60/007,473	22 November 1995 (22.11.95)	US
60/008,992	21 December 1995 (21.12.95)	US
60/013,748	20 March 1996 (20.03.96)	US
60/013,764	20 March 1996 (20.03.96)	US
60/013,747	20 March 1996 (20.03.96)	US
60/017,455	17 May 1996 (17.05.96)	US
60/017,892	17 May 1996 (17.05.96)	US
60/020,478	13 June 1996 (13.06.96)	US
60/022,047	22 July 1996 (22.07.96)	US
60/023,494	7 August 1996 (07.08.96)	US
60/023,742	8 August 1996 (08.08.96)	US

(60) Parent Applications or Grants
(63) Related by Continuation

US	60/008,108 (CIP)
Filed on	30 October 1995 (30.10.95)
US	60/007,473 (CIP)
Filed on	22 November 1995 (22.11.95)
US	60/008,992 (CIP)
Filed on	21 December 1995 (21.12.95)
US	60/013,748 (CIP)
Filed on	20 March 1996 (20.03.96)
US	60/013,764 (CIP)
Filed on	20 March 1996 (20.03.96)
US	60/013,747 (CIP)
Filed on	20 March 1996 (20.03.96)
US	60/017,892 (CIP)
Filed on	17 May 1996 (17.05.96)
US	60/017,455 (CIP)
Filed on	17 May 1996 (17.05.96)
US	60/020,478 (CIP)
Filed on	13 June 1996 (13.06.96)
US	60/022,047 (CIP)
Filed on	22 July 1996 (22.07.96)
US	60/023,494 (CIP)
Filed on	7 August 1996 (07.08.96)
US	60/023,742 (CIP)
Filed on	8 August 1996 (08.08.96)

(71) Applicant (*for all designated States except US*): SMITHKLINE BEECHAM CORPORATION [US/US]; Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

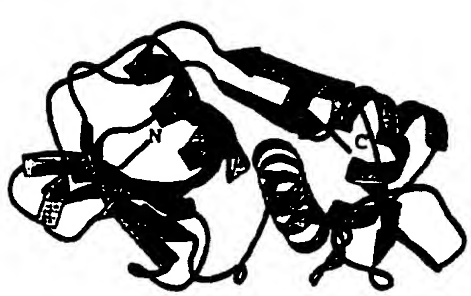
(72) Inventors; and
(75) Inventors/Applicants (*for US only*): ABDEL-MEQUID, Sherin, Salaheldin [US/US]; 236 Autumn Drive, Exton, PA 19341 (US). CARR, Thomas, Joseph [US/US]; 27 Jonathan Drive, Phoenixville, PA 19460 (US). DESJARLAIS, Renee, Louise [US/US]; 11 Cornwall Circle, St. Davids, PA 19087 (US). GALLAGHER, Thimothy, Francis [US/US]; 255 Manor Road, Harleysville, PA 19438 (US). HALBERT, Stacie, Marie [US/US]; 149 Montgomery Drive, Harleysville, PA 19438 (US). JANSON, Cheryl, Ann [US/US]; 200 Ladbroke Road, Bryn Mawr, PA 19010 (US). MARQUIS, Robert, Wells, Jr. [US/US]; 115 Cambria Court, St. Davids, PA 19087 (US). OH, Hye-Ja [US/US]; 326 Long Ridge Lane, Exton, PA 19341 (US). RU, Yu [US/US]; 109 Gilmore Road, Havertown, PA 19083 (US). SMITH, Ward, Whitlock, Jr. [US/US]; 200 Ladbroke Road, Bryn Mawr, PA 19010 (US). THOMPSON, Scott, Kevin [US/US]; 75 Guilford Circle, Phoenixville, PA 19460 (US). VEBER, Daniel, Frank [US/US]; 290 Batleson Road, Ambler, PA 19002 (US). YAMASHITA, Dennis, Shinji [US/US]; 703 Edgewood Road, King of Prussia, PA 19406 (US). YEN, Jack, Hwekwo [US/US]; 273 Phoenixville Pike, Malvern, PA 19355 (US). ZHAO, Baoguang [CN/US]; 649 South Henderson Road, A312, King of Prussia, PA 19406 (US).

(74) Agents: McCARTHY, Mary, E. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

(81) Designated States: AL, AM, AU, BB, BG, BR, CA, CN, CZ, EE, GE, HU, IL, IS, JP, KG, KP, KR, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD OF INHIBITING CATHEPSIN K



Human Cathepsin K

(57) Abstract

A novel cathepsin K crystalline structure is identified. Also disclosed are methods of identifying inhibitors of this protease and methods of inhibiting cathepsin K using inhibitors with certain structural, physical and spatial characteristics.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

METHOD OF INHIBITING CATHEPSIN K

Field of the Invention

This invention relates to a method of inhibiting cathepsin K by administering compounds with certain structural, physical and spatial characteristics that allow for the interaction of said compounds with specific residues of the active site of the enzyme. This interaction between the compounds of this invention and the active site inhibits the activity of cathepsin K and these compounds are useful for treating diseases in which said inhibition is indicated, such as osteoporosis and periodontal disease. This invention also relates to a novel crystalline structure of cathepsin K, the identification of a novel protease catalytic active site for this enzyme and methods enabling the design and selection of inhibitors of said active site.

Background of the Invention

Cathepsin K is a member of the family of enzymes which are part of the papain superfamily of cysteine proteases. Cathepsins B, H, L, N and S have been described in the literature. Recently, cathepsin K polypeptide and the cDNA encoding such polypeptide were disclosed in U.S. Patent No. 5,501,969 (called cathepsin O therein). Cathepsin K has been recently expressed, purified, and characterized. Bossard, M. J., et al., (1996) *J. Biol. Chem.* **271**, 12517-12524; Drake, F.H., et al., (1996) *J. Biol. Chem.* **271**, 12511-12516; Bromme, D., et al., (1996) *J. Biol. Chem.* **271**, 2126-2132.

Cathepsin K has been variously denoted as cathepsin O, cathepsin X or cathepsin O2 in the literature. The designation cathepsin K is considered to be the more appropriate one (name assigned by Nomenclature Committee of the International Union of Biochemistry and Molecular Biology).

Cathepsins of the papain superfamily of cysteine proteases function in the normal physiological process of protein degradation in animals, including humans, e.g., in the degradation of connective tissue. However, elevated levels of these enzymes in the body can result in pathological conditions leading to disease. Thus, cathepsins have been implicated in various disease states, including but not limited to, infections by pneumocystis carinii, trypsanoma cruzi, trypsanoma brucei brucei, and Crithidia fusiculata; as well as in schistosomiasis malaria, tumor metastasis, metachromatic leukodystrophy, muscular dystrophy, amytrophy, and the like. *See* International Publication Number WO 94/04172, published on March 3, 1994, and

references cited therein. *See also* European Patent Application EP 0 603 873 A1, and references cited therein. Two bacterial cysteine proteases from P. gingivallis, called gingipains, have been implicated in the pathogenesis of gingivitis. Potempa, J., et al. (1994) *Perspectives in Drug Discovery and Design*, 2, 445-458.

Cathepsin K is believed to play a causative role in diseases of excessive bone or cartilage loss. Bone is composed of a protein matrix in which spindle- or plate-shaped crystals of hydroxyapatite are incorporated. Type I Collagen represents the major structural protein of bone comprising approximately 90% of the structural protein. The remaining 10% of matrix is composed of a number of non-collagenous proteins, including osteocalcin, proteoglycans, osteopontin, osteonectin, thrombospondin, fibronectin, and bone sialoprotein. Skeletal bone undergoes remodeling at discrete foci throughout life. These foci, or remodeling units, undergo a cycle consisting of a bone resorption phase followed by a phase of bone replacement.

Bone resorption is carried out by osteoclasts, which are multinuclear cells of hematopoietic lineage. The osteoclasts adhere to the bone surface and form a tight sealing zone, followed by extensive membrane ruffling on their apical (i.e., resorbing) surface. This creates an enclosed extracellular compartment on the bone surface that is acidified by proton pumps in the ruffled membrane, and into which the osteoclast secretes proteolytic enzymes. The low pH of the compartment dissolves hydroxyapatite crystals at the bone surface, while the proteolytic enzymes digest the protein matrix. In this way, a resorption lacuna, or pit, is formed. At the end of this phase of the cycle, osteoblasts lay down a new protein matrix that is subsequently mineralized. In several disease states, such as osteoporosis and Paget's disease, the normal balance between bone resorption and formation is disrupted, and there is a net loss of bone at each cycle. Ultimately, this leads to weakening of the bone and may result in increased fracture risk with minimal trauma.

The abundant selective expression of cathepsin K in osteoclasts strongly suggests that this enzyme is essential for bone resorption. Thus, selective inhibition of cathepsin K may provide an effective treatment for diseases of excessive bone loss, including, but not limited to, osteoporosis, gingival diseases such as gingivitis and periodontitis, Paget's disease, hypercalcemia of malignancy, and metabolic bone disease. Cathepsin K levels have also been demonstrated to be elevated in chondroclasts of osteoarthritic synovium. Thus, selective inhibition of cathepsin K may also be useful for treating diseases of excessive cartilage or matrix degradation.

including, but not limited to, osteoarthritis and rheumatoid arthritis. Metastatic neoplastic cells also typically express high levels of proteolytic enzymes that degrade the surrounding matrix. Thus, selective inhibition of cathepsin K may also be useful for treating certain neoplastic diseases.

Surprisingly, it has been found that a broad, structurally diverse series of compounds have common structural, physical and spatial characteristics that allow for the interaction of said compounds with specific residues of the active site of cathepsin K and are useful for treating diseases in which inhibition of bone resorption is indicated, such as osteoporosis and periodontal disease. Thus, this invention relates to the method of inhibiting cathepsin K using compounds having the characteristics hereinbelow defined.

Summary of the Invention

In one aspect, the present invention provides a method for inhibiting cathepsin K by administering compounds with certain structural, physical and spatial characteristics that allow for the interaction of said compounds with specific residues of the active site of the enzyme. This interaction inhibits the activity of cathepsin K and, thus, treats diseases in which bone resorption is a factor.

In another aspect, the present invention provides a novel cysteine protease in crystalline form.

In yet another aspect, the invention provides a novel protease composition characterized by a three dimensional catalytic site formed by the atoms of the amino acid residues listed in Table XXIX.

In still another aspect, the invention provides a method for identifying inhibitors of the compositions described above which methods involve the steps of: providing the coordinates of the protease structure of the invention to a computerized modeling system; identifying compounds which will bind to the structure; and screening the compounds or analogs derived therefrom identified for cathepsin K inhibitory bioactivity.

Other aspects and advantages of the present invention are described further in the following detailed description of the preferred embodiments thereof.

Brief Description of the Drawings

Figure 1 is the amino acid sequence of cathepsin K aligned with the amino acid sequences of other cysteine proteases.

Figure 2 is a ribbon diagram of cathepsin K. The amino and carboxyl-termini are indicated by N and C. The drawing was produced using the program MOLSCRIPT [Kraulis, P., *J. Appl. Crystallogr.*, **24**, 946-950 (1991)].

Figure 3 is a ribbon diagram of cathepsin K in complex with E-64, a known inhibitor of cysteine proteases. The drawing was produced using the program MOLSCRIPT.

Figure 4a is an illustration of the active site of cathepsin K. Figure 4b is a stereoview of the active site of cathepsin K. For clarity, no hydrogen atoms or water molecules are shown.

Figures 5a-13a are illustrations of the active site of cathepsin K in complex with novel inhibitors of cathepsin K. Figures 5b-13b are stereoviews of the active site of cathepsin K in complex with novel inhibitors of cathepsin K. These views depict the interaction of each inhibitor with all atoms of residues of the active site of cathepsin K within 5Å of the inhibitors. For clarity, no hydrogen atoms or water molecules are shown.

Table I provides the three dimensional protein coordinates of the cathepsin K crystalline structure of the invention.

Tables II-X provide the three dimensional coordinates for the cathepsin K complex with specific inhibitors of the present invention.

Tables XI-XIX provide listings of the three atom angles between atoms of the inhibitors and the protein for all inhibitor atoms within 5 Ångstroms of the protein.

Tables XX-XXVIII provide listings of the distances between atoms of the inhibitors and the protein for all inhibitor atoms within 5 Ångstroms of the protein.

Table XXIX provides the atoms of the amino acid residues of the catalytic site.

Detailed Description of the Invention

The present invention provides a novel cysteine protease crystalline structure, a novel cysteine protease active site, and methods of use of the crystalline form and active site to identify protease inhibitor compounds.

In particular, the present invention provides a method for inhibiting cathepsin K by administering compounds with certain structural, physical and spatial characteristics that allow for the interaction of said compounds with specific residues

Specifically, the inhibitors of cathepsin K used in the present invention interact with any two or more of the following:

1. Tyrosine 67 sidechain;
2. Hydrophobic pocket lined with atoms from methinoine 68, leucine 209, alanine 163, alanine 134 and portions of tyrosine 67;
3. Hydrogen bonds donated by glycine 66 amide nitrogen;
4. Cysteine 25 the active site nucleophile;
5. Mainchain interactions from residues glutamine 21, cysteine 22, and glycine 23;
6. Tryptophan 184 sidechain; and
7. Hydrophobic contacts with the sidechain atoms of glutamine 143 and asparagine 161 and the mainchain of alanine 137 and serine 138.

Preferably, the inhibitors of cathepsin K used in the present invention interact with any three or more of the above-identified regions of the active site.

The compounds used in the methods of the present invention possess an electrophilic carbon and either a hydrophobic group whose centroid is 5.44-6.94Å from the carbon or an aromatic group whose centroid is 9.24-11.24Å from the carbon, or both the hydrophobic and the aromatic groups in which case the centroids of these two groups should be 15.67-16.67Å apart. These features must be able to make the appropriate interactions with the cathepsin K active site. The electrophilic carbon atom should be 1.7-4.0Å from the side chain sulfur atom (SG) on the amino acid cysteine 25. The hydrophobic group should be near the following amino acids with appropriate distance ranges between the centroid of the side chain atoms and the centroid of the hydrophobic group given in parentheses: tyrosine 67 (4.91-5.91Å), methionine 68 (5.74-6.74Å), alanine 134 (4.15-5.15Å), leucine 160 (6.18-7.18Å), and leucine 209 (5.71-6.71Å). The aromatic group should be near the either tryptophan 184 (4.10-7.10Å) or tryptophan 188 (4.10-7.10Å) or both.

The key structural features of the inhibitors of the present invention include an electrophilic carbon, preferably the carbon of a carbonyl group, a hydrophobic group, preferably an isobutyl group, and an aromatic group, preferably a phenyl group. The electrophilic carbon of the inhibitor may be in the same compound with two hydrophobic groups, such as two isobutyl groups, or two aromatic groups, such as two phenyl groups, or one hydrophobic group and one aromatic group.

Suitably, the method of inhibiting cathepsin K of the present invention comprises administering to a mammal, preferably a human, in need thereof a

compound that fits spatially into the active site of cathepsin K, said compound comprising any two or more of the following:

(i) an electrophilic carbon atom that binds to the side chain sulfur atom of cysteine 25 wherein said electrophilic carbon atom is 1.7-4.0Å from said sulfur atom;

(ii) a hydrophobic group that interacts with tryptophan 184 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tryptophan 184 is 4.10-7.10Å;

(iii) a hydrophobic group that interacts with tyrosine 67, methionine 68, alanine 134, leucine 160, and leucine 209, creating a hydrophobic pocket, and has distance ranges between the centroid of said hydrophobic group and the centroids of the side chain atoms of the amino acid residues of said hydrophobic pocket which are tyrosine 67: 4.91- 5.91Å, methionine 68: 5.74-6.74Å, alanine 134: 4.15-5.15Å, leucine 160: 6.18-7.18Å, and leucine 209: 5.71-6.71Å;

(iv) a hydrophobic group that interacts with tyrosine 67 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tyrosine 67 is 4.10-7.10Å;

(v) an amino group with a pKa of less than 7 or an oxygen atom, each of which interacts with a hydrogen atom donated by the amide nitrogen of glycine 66 wherein the distance between these two atoms is 2.7-3.5Å;

(vi) a hydrophobic group that interacts with the main chain atoms of glutamine 21, cysteine 22 and glycine 23 wherein the distance between the centroid of said hydrophobic group and the centroids of glutamine 21 , cysteine 22 and glycine 23 are 3.7-5.4, 4.9-5.7 and 5.4-6.7Å, respectively; or

(vii) a hydrophobic group that interacts with the side chain atoms of glutamine 143 and asparagine 161 and the main chain of alanine 137 and serine 138 wherein the distance between the centroid of the hydrophobic group and the centroids of glutamine 143, asparagine 161, alanine 137, and serine 138 are 7.9-9.6Å, 4.7-5.4Å, 4.2-5.5Å, and 4.6-6.4Å, respectively. Preferably, the inhibitors of cathepsin K used in the present invention comprise three or more of the above.

Suitably, the method of inhibiting cathepsin K of the present invention comprises administering to a mammal, preferably a human, in need thereof, a compound that fits spatially into the active site of cathepsin K, said compound comprising:

(i) an electrophilic carbon atom that binds to the side chain sulfur atom of cysteine 25 wherein said electrophilic carbon atom is 1.7-4.0Å from said sulfur atom; and

(ii) a hydrophobic group that interacts with tryptophan 184 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tryptophan 184 is 4.10-7.10Å. Preferably, the hydrophobic group that interacts with tryptophan 184 is an aromatic group and the centroid of this aromatic group is 9.24-11.24Å from the centroid of the electrophilic carbon that binds to the side chain sulfur atom of cysteine 25.

Preferably, the electrophilic carbon that binds to the side chain sulfur atom of cysteine 25 is a carbonyl carbon.

Suitably, the method of the present invention further comprises a compound with a hydrophobic group that:

has a centroid which is 5.44-6.94Å from said electrophilic carbon;

interacts with tyrosine 67, methionine 68, alanine 134, leucine 160, and leucine 209, creating a hydrophobic pocket; and

has distance ranges between the centroid of said hydrophobic group and the centroids of the side chain atoms of the amino acid residues of said hydrophobic pocket which are tyrosine 67: 4.91- 5.91Å, methionine 68: 5.74-6.74Å, alanine 134: 4.15-5.15Å, leucine 160: 6.18-7.18Å, and leucine 209: 5.71-6.71Å. Preferably, this hydrophobic group is an isobutyl group.

Alternately, the method of the present invention further comprises a compound with a hydrophobic group that interacts with tyrosine 67 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tyrosine 67 is 4.10-7.10Å. Preferably, this hydrophobic group is an aromatic group.

Alternately, the method of the present invention further comprises a compound with an amino group with a pKa of less than 7 or an oxygen atom, each of which interacts with a hydrogen atom donated by the amide nitrogen of glycine 66 wherein the distance between these two atoms is 2.7-3.5Å. Preferably, the compound comprises an oxygen atom, such as an oxygen atom of a carbonyl group or an oxygen atom of a hydroxyl group.

Alternately, the method of the present invention further comprises a compound with a hydrophobic group that interacts with the main chain atoms of glutamine 21, cysteine 22 and glycine 23 wherein the distance between the centroid

of the hydrophobic group and the centroids of glutamine 21, cysteine 22 and glycine 23 are 3.7-5.4, 4.9-5.7 and 5.4-6.7Å, respectively. Preferably, this hydrophobic group is an isobutyl group.

Alternately, the method of the present invention further comprises a compound with a hydrophobic group that interacts with the side chain atoms of glutamine 143 and asparagine 161 and the mainchain of alanine 137 and serine 138 wherein the distance between the centroid of the hydrophobic group and the centroids of glutamine 143, asparagine 161, alanine 137, and serine 138 are 7.9-9.6Å, 4.7-5.4Å, 4.2-5.5Å, and 4.6-6.4Å, respectively.

Compounds used in the method of the present invention include, but are not limited to, the following:

3(S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone;

4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone;

4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N-[N-(methyl)-L-leucyl)]-3-pyrrolidinone;

4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone;

bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one;

2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide;

(1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide;

1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one; and

2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide;

or a pharmaceutically acceptable salt thereof.

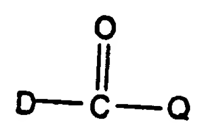
As stated herein, the interaction of the inhibitor at the side chain sulfur atom of cysteine 25 has as one of its requirements that the inhibitor contain an "electrophilic carbon" atom. By this term is meant an electron deficient carbon. This term includes, but is not limited to, a carbonyl carbon atom. This term also includes an epoxide, a thiocarbonyl, an imine, and a nitrile. Suitably, this term may also be represented by the formula -C=N-X, wherein X may be optionally tied back to C in a ring or wherein X is CH_2, H, O, S or NR^a in which R^a is H of C_{1-4}alkyl.

includes an epoxide, a thiocarbonyl, an imine, and a nitrile. Suitably, this term may also be represented by the formula -C=N-X, wherein X may be optionally tied back to C in a ring or wherein X is CH_2, H, O, S or NR^a in which R^a is H of C_{1-4}alkyl.

The hydrophobic groups that interact with tryptophan 184 or tyrosine 67 include, but are not limited to, aromatic groups. These hydrophobic groups include phenyl, C_{1-6}alkyl and heteroaryl, which is defined hereinbelow. The hydrophobic groups that interact with the hydrophobic pocket lined with atoms from tyrosine 67, methionine 68, alanine 134, leucine 160, and leucine 209 not only includes isobutyl, but also includes C_{1-6}alkyl, C_{3-6}cycloalkyl and adamantyl. The hydrophobic groups that interact with the main chain atoms of glutamine 21, cysteine 22 and glycine 23 or the side chain atoms of glutamine 143 and asparagine 161 and the mainchain of alanine 137 and serine 138 include C_{1-10}alkyl, C_bF_{2b+1}, in which b is 1-3, and aryl and heteroaryl, each of which are defined hereinbelow.

As used herein, the term "centroid" means the position for the stated atoms calculated by averaging the x coordinates of the atoms to obtain the x coordinate of the centroid, averaging the y coordinates of the atoms to obtain the y coordinate of the centroid, and averaging the z coordinates of the atoms to obtain the z coordinate of the centroid.

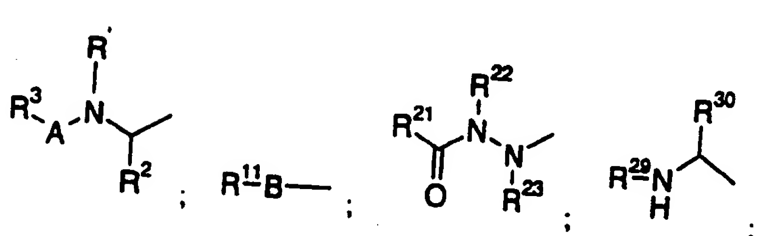
The compounds used in the method of the present invention include, but are not limited to, the compounds of formula (I):

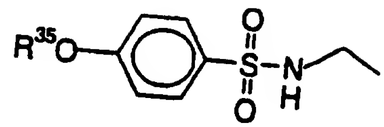


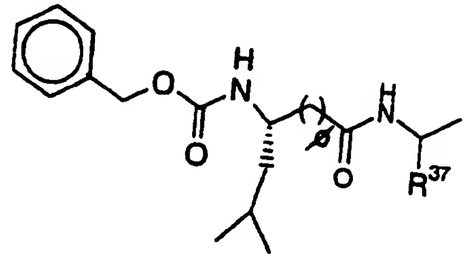
I

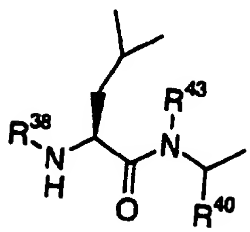
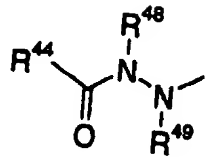
wherein:

D =

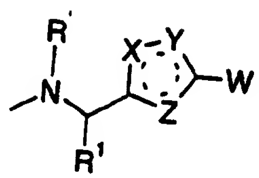
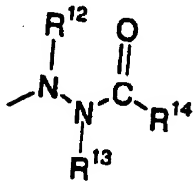
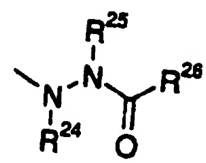


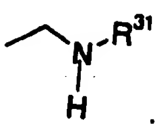
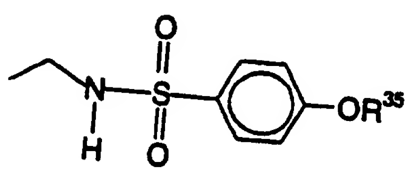
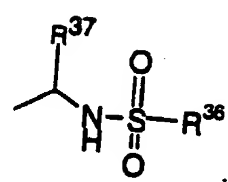
 ;

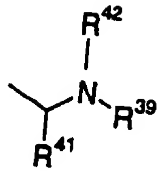
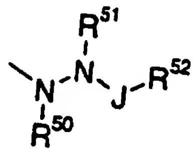
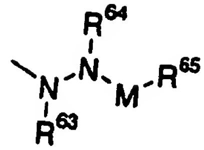
 ;

 ;  ; L—G— ;

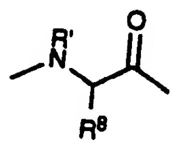
Q =

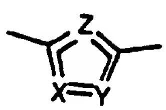
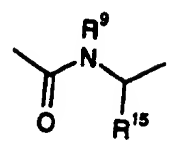
 ;  ; 

 ;  ;  ;

 ;  ;  ;

where:

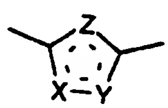
A = absent,  ;

B =  ,  ;

L = C_{2-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl, CH(R^{66})N$R^{60}R^{68}$, CH(R^{66})Ar, CH(R^{66})OAr', N$R^{66}R^{67}$;

M = C(O), SO_2;

G =



J = C(O), SO_2;

T = Ar, Het;

V = C_{3-7}cycloalkyl;

W = H, -CN, -CF_3, -NO_2, -COR^7, -CO_2R^6, -CONHR^6, -SO_2NHR^6, -NHSO_2R^6, -NHCOR^7, -O-COR^6, -SR^6, NR'R^6, NR'(C=NH)NHR^5, Cl, Br, I, F;

$X = Y = Z$ = N, O, S or CR^4,

provided that at least two of X, Y and Z are heteroatoms and at least one of X, Y and Z is N, or one of X, Y and Z is C=N, C=C or N=N and the other two are CR^4 or N, provided that X, Y and Z together comprise at least two N;

$=$ indicates a single or double bond in the five-membered heterocycle;

m = 0, 1, 2;

n = 1 to 6;

f = 0, 1, 2;

Ar = phenyl, naphthyl, optionally substituted by one or more of Ph-C_{0-6}alkyl, Het-C_{0-6}alkyl, C_{1-6}alkoxy, Ph-C_{0-6}alkoxy, Het-C_{0-6}alkoxy, OH, $(CH_2)_{1-6}NR^{58}R^{59}$, $O(CH_2)_{1-6}NR^{58}R^{59}$;

Ar' = phenyl or naphthyl, optionally substituted by one or more of Ph-C_{0-6}alkyl, Het-C_{0-6}alkyl, C_{1-6}alkoxy, Ph-C_{0-6}alkoxy, Het-C_{0-6}alkoxy, OH, $(CH_2)_{1-6}NR^{58}R^{59}$, $O(CH_2)_{1-6}NR^{58}R^{59}$, or halogen;

R' = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^1 = H, C_{1-6}alkyl;

R^2 = C_{4-6}alkyl, C_{4-6}alkenyl, benzyl;

R^3 = C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl, R^5CO-, R^5SO_2-, R^5OC(O)-, R^5NHCO-;

R^4 = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^5 = Ar-0-6alkyl, Het-C_{0-6}alkyl;

R^6 = H, C_{1-6}alkyl, CH_2CF_3, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

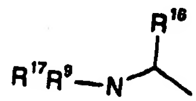
R^7 = C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^8 = H; C_{2-6} alkenyl; C_{2-6}alkynyl; Het; Ar; C_{1-6}alkyl, optionally substituted by OR', SR', NR'_2, CO_2R', $CO_2NR'_2$, $N(C{=}NH)NH_2$, Het or Ar;

R^9 = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^{10} = C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

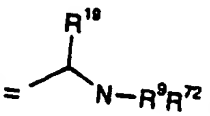
R^{11} = H, C_{1-6}alkyl, Ar-C_{1-6}alkyl, Het-C_{0-6}alkyl, or



;

R^{12} = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^{13} = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^{14} = 

, Ac;

R^{15} = H, C_{1-6}alkyl, C_{2-6}alkenyl, C_{2-6}alkynyl, Ar, Het, or C_{1-6}alkyl optionally substituted by OR^9, NR^9_2, $CONR^9_2$, N(C=NH)NH-, Het or Ar;

R^{16} = C_{2-6}alkyl, C_{2-6}alkenyl, C_{2-6}alkynyl, Ar, Het, or C_{2-6}alkyl optionally substituted by OR^9, SR^9, NR^9_2, CO_2R^9, $CONR^9_2$, N(C=NH)NH-, Het or Ar;

R^{19} = H, C_{1-6}alkyl, C_{2-6}alkenyl, C_{2-6}alkynyl, Ar, Het, or C_{1-6}alkyl optionally substituted by OR^9, SR^9, NR^9_2, CO_2R^9, $CONR^9_2$, N(C=NH)NH-, Het or Ar;

R^{17} = R^{72} = H, C_{1-6}alkyl, R^{10}, R^{10}C(O)-, R^{10}C(S)-, R^{10}OC(O)-;

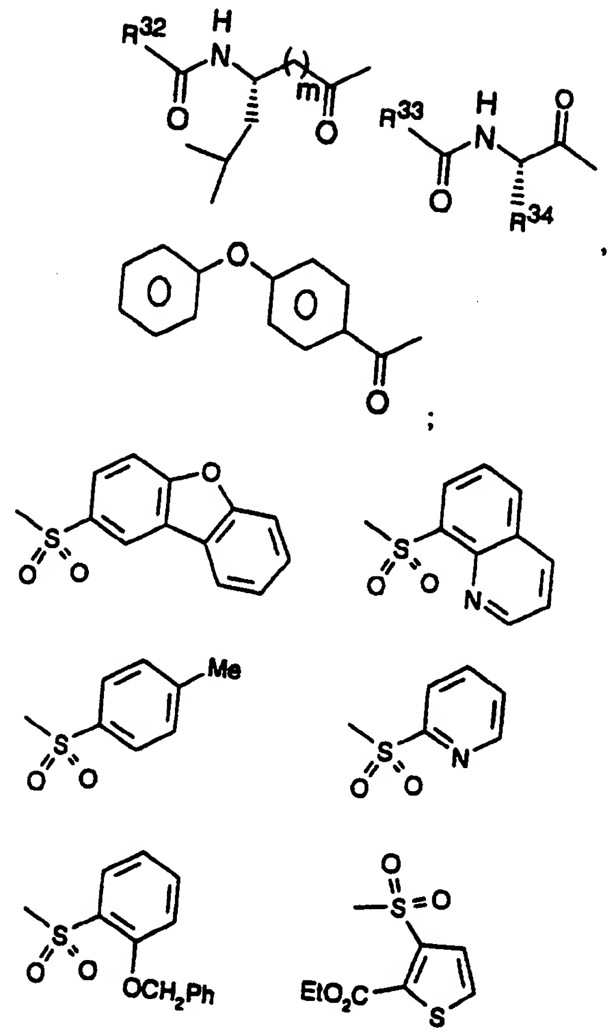
$R^{21} = R^{26} = C_{5-6}$alkyl; C_{2-6}alkenyl; C_{3-11}cycloalkyl; T-C_{3-6}alkyl; V-C_{1-6}alkyl; T-C_{2-6}alkenyl; T- $(CH_2)_nCH(T)(CH_2)_n$; optionally substituted by one or two halogens, SR^{20}, OR^{20},$NR^{20}R^{27}$ or C_{1-4}alkyl;

$R^{27} = R^{28}CO$, $R^{28}OCO$;

$R^{28} = C_{1-6}$alkyl; C_{3-11}cycloalkyl; Ar; Het; T-C_{1-6}alkyl; T-$(CH_2)_nCH(T)(CH_2)_n$; optionally substituted by one or two halogens, SR^{20}, OR^{20}, $NR^{20}R^{73}$, C_{1-6}alkyl;

$R^{20} = R^{22} = R^{23} = R^{24} = R^{25} = R^{73} =$ H, C_{1-4}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

$R^{29} =$

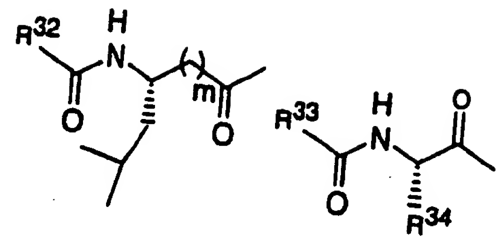


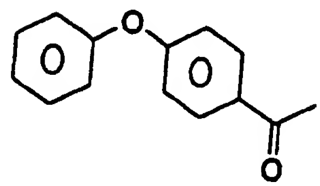
Cbz-leucinyl-; 2-, 3-, or 4-pyridyl methyloxycarbonyl-leucinyl-; 4-imidazole acetyl-leucinyl-, phenyl acetyl-leucinyl, N,N-dimethyl-glycinyl leucinyl, 4-pyridyl acetyl-leucinyl, 2-pyridyl sulfonyl-leucinyl, 4-pyridyl carbonyl-leucinyl, acetyl-leucinyl, benzoyl-leucinyl, 4-phenoxy-benzoyl-, 2- or 3-benzyloxybenzoyl-, biphenyl acetyl, lpha- isobutyl-biphenyl acetyl, Cbz-phenylalaninyl, Cbz-norleucinyl-, Cbz-norvalinyl-, Cbz-glutamyl-, Cbz-

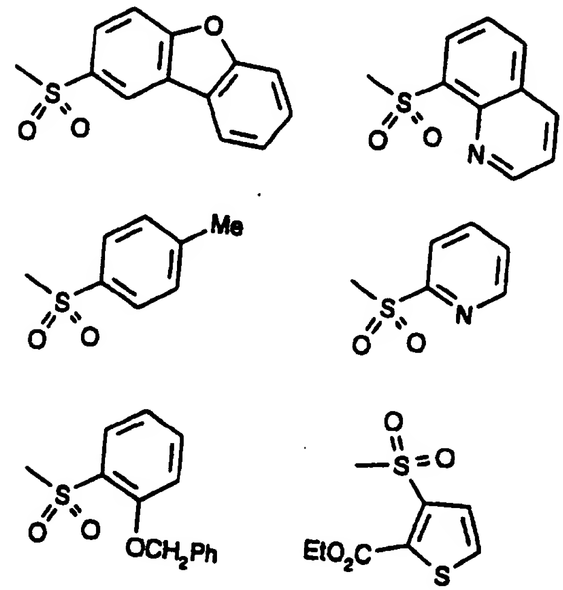
epsilon- (t-butyl ester)-glutamyl; acetyl-leucinyl-, 6- or 8- quinoline carbonyl, biphenyl acetyl, alpha- isobutyl-biphenyl acetyl, acetyl, benzoyl, 2- or 3- benzyloxy benzoyl, 4-phenoxy benzoyl-, Cbz-amino acid-; 2-,3-, or 4-pyridylmethyloxycarbonyl-aminoacid-; aryl C_0-C_6alkyloxy carbonyl-amino acid-, heteroaryl C_0-C_6alkyloxy carbonyl-amino acid-,aryl C_0-C_6alkyloxy carbonyl-amino acid-, heteroaryl C_0-C_6alkyloxy carbonyl-amino acid-, C_1-C_6alkyloxy carbonyl-amino acid-; C_1-C_6alkyl carbonyl, aryl C_0-C_6alkyl carbonyl, heteroaryl C_0-C_6alkyl carbonyl, aryl C_0-C_6alkyl carbonyl, heteroaryl C_0-C_6alkyl carbonyl, C_1-C_6alkyl sulfonyl, aryl C_0-C_6alkyl sulfonyl, heteroarylC_0-C_6alkyl sulfonyl, aryl C_0-C_6alkyl sulfonyl, heteroaryl C_0-C_6alkyl sulfonyl;

R^{30} = -H, C_{1-6} alkyl;

R^{31} =


,


;



Cbz-leucinyl-; 2-, 3-, or 4-pyridyl methyloxycarbonyl-leucinyl-; 4-imidazole acetyl-leucinyl-, phenyl acetyl-leucinyl, N,N-dimethyl-glycinyl leucinyl, 4-pyridyl acetyl-leucinyl, 2-pyridyl sulfonyl-leucinyl, 4-pyridyl carbonyl-leucinyl, acetyl-leucinyl, benzoyl-leucinyl, 4-phenoxy-benzoyl-, 2- or 3-benzyloxybenzoyl-, biphenyl acetyl, alpha- isobutyl-biphenyl acetyl, Cbz-phenylalaninyl, Cbz-norleucinyl-, Cbz-norvalinyl-, Cbz-glutamyl-, Cbz-

epsilon- (t-butyl ester)-glutamyl; acetyl-leucinyl-, 6- or 8- quinoline carbonyl, biphenyl acetyl, alpha- isobutyl-biphenyl acetyl, acetyl, benzoyl, 2- or 3- benzyloxy benzoyl, 4-phenoxy benzoyl-, Cbz-amino acid-; 2-,3-, or 4-pyridylmethyloxycarbonyl-aminoacid-; aryl C_0-C_6alkyloxy carbonyl-amino acid-, heteroaryl C_0-C_6alkyloxy carbonyl-amino acid-,aryl C_0-C_6alkyloxy carbonyl-amino acid-, heteroaryl C_0-C_6alkyloxy carbonyl-amino acid-, C_1-C_6alkyloxy carbonyl-amino acid-; C_1-C_6alkyl carbonyl, aryl C_0-C_6alkyl carbonyl, heteroaryl C_0-C_6alkyl carbonyl, aryl C_0-C_6alkyl carbonyl, heteroaryl C_0-C_6alkyl carbonyl, C_1-C_6alkyl sulfonyl, aryl C_0-C_6alkyl sulfonyl, heteroaryl C_0-C_6alkyl sulfonyl, aryl C_0-C_6alkyl sulfonyl, heteroaryl C_0-C_6alkyl sulfonyl;

R^{32} = OCH_2Ar, OCH2$C_{1\text{-}6}$alkyl, aryl substituted $C_{0\text{-}6}$alkyl, heteroaryl substituted $C_{0\text{-}6}$alkyl,4-imidazole methylene; 2-, 3-, or 4-pyridylmethylneneoxy; 4-pyridyl methylene, 2-pyridyl sulfonyl, 4-pyridyl, aryl substituted $C_{0\text{-}6}$alkyloxy, heteroaryl substituted $C_{0\text{-}6}$alkyloxy;

R^{33} = $C_{1\text{-}6}$alkyl, $-CH_2Ph$, $-CH_2CH_2CO_2R^{34}$;

R^{34} = -H, $C_{1\text{-}6}$alkyl;

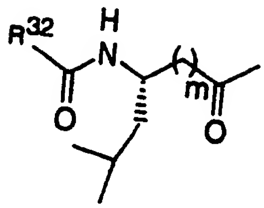
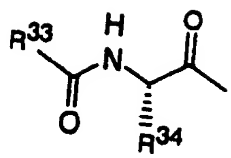
R^{35} = Ar, HetAr;

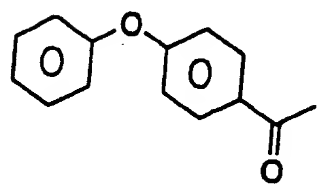
R^{36} = Aryl, heteroaryl, pyridyl, isoquinolinyl;

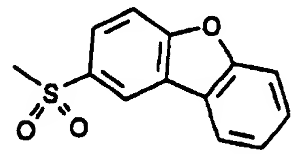
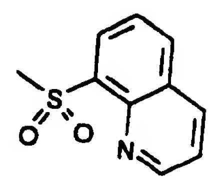
R^{37}= $C_{1\text{-}6}$alkyl, $-CH_2Ph$, $-CH_2CH_2CO_2R^{34}$;

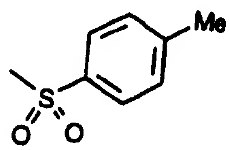
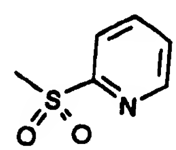
R^{38} = Cbz; $C_{1\text{-}6}$alkyl or aryl substituted Cbz; $C_{1\text{-}6}$alkyl -CO; benzoyl; $C_{1\text{-}6}$alkyl or aryl substituted benzoyl;

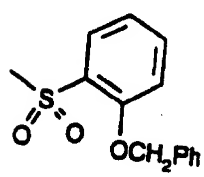
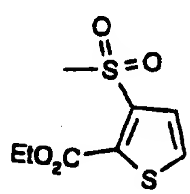
R^{39} =

  ,

 ;

Cbz-leucinyl-; 2-, 3-, or 4-pyridyl methyloxycarbonyl-leucinyl-; 4-imidazole acetyl-leucinyl-, phenyl acetyl-leucinyl, N,N-dimethyl-glycinyl leucinyl, 4-pyridyl acetyl-leucinyl, 2-pyridyl sulfonyl-leucinyl, 4-pyridyl carbonyl-leucinyl, acetyl-leucinyl, benzoyl-leucinyl, 4-phenoxy-benzoyl-, 2- or 3-benzyloxybenzoyl-, biphenyl acetyl, alpha- isobutyl-biphenyl acetyl, Cbz-phenylalaninyl, Cbz-norleucinyl-, Cbz-norvalinyl-, Cbz-glutamyl-, Cbz-

epsilon- (t-butyl ester)-glutamyl; acetyl-leucinyl-, 6- or 8- quinoline carbonyl, biphenyl acetyl, alpha- isobutyl-biphenyl acetyl, acetyl, benzoyl, 2- or 3- benzyloxy benzoyl, 4-phenoxy benzoyl-, Cbz-amino acid-; 2-,3-, or 4- pyridylmethyloxycarbonyl-aminoacid-; aryl C_0-C_6alkyloxy carbonyl-amino acid-, heteroaryl C_0-C_6alkyloxy carbonyl-amino acid-,aryl C_0-C_6alkyloxy carbonyl-amino acid-,heteroaryl C_0-C_6alkyloxy carbonyl-amino acid-, C_1-C_6alkyloxy carbonyl-amino acid-; C_1-C_6alkyl carbonyl, aryl C_0-C_6alkyl carbonyl, heteroaryl C_0-C_6alkyl carbonyl, aryl C_0-C_6alkyl carbonyl, heteroaryl C_0-C_6alkyl carbonyl, C_1-C_6alkyl sulfonyl, aryl C_0-C_6alkyl sulfonyl, heteroaryl C_0-C_6alkyl sulfonyl, aryl C_0-C_6alkyl sulfonyl, heteroaryl C_0-C_6alkyl sulfonyl;

R^{40} = H and C_{1-6}alkyl;

R^{41} = H and C_{1-6}alkyl;

R^{42} = C_{1-6}alkyl, aryl substituted C_{1-6}alkyl and hetero aryl substituted C_{1-6}alkyl,; H when R^{43} is C_{1-6}alkyl, aryl substituted C_{1-6}alkyl; and heteroaryl substituted C_{1-6}alkyl;

R^{43} = C_{1-6}alkyl, aryl substituted C_{1-6}alkyl and hetero aryl substituted C_{1-6}alkyl,; H when R^{42} is C_{1-6}alkyl, aryl substituted C_{1-6}alkyl; and heteroaryl substituted C_{1-6}alkyl;

R^{44} = $CH(R^{53})NR^{45}R^{54}$, $CH(R^{55})Ar$, C_{5-6}alkyl;

$R^{45} = R^{46} = R^{47} = R^{48} = R^{49} = R^{50} = R^{51}$ = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^{52} = Ar, Het, $CH(R^{56})Ar$, $CH(R^{56})OAr$, $N(R^{56})Ar$, C_{1-6}alkyl, $CH(R^{56})NR^{46}R^{57}$;

R^{53} = C_{2-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl,

R^{53} and R^{45} may be connected to form a pyrrolidine or piperidine ring;

R^{54} = R^{57} = R^{47}, R^{47}C(O), R^{47}C(S), R^{47}OC(O);

R^{55} = R^{56} = R^{58} = R^{59} = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^{60} = R^{61} = R^{62} = R^{63} = R^{64} = H, C_{1-6}alkyl, Ar-C_{0-6}alkyl, or Het-C_{0-6}alkyl;

R^{65} = C_{1-6}alkyl, Ar, Het, CH(R^{69})Ar, CH(R^{69})OAr, N(R^{69})Ar, CH(R^{69})N$R^{61}R^{70}$;

R^{66} = R^{69} = R^{71} = H, C_{1-6}alkyl, $(CH_2)_{0-6}$-C_{3-6}cycloalkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl;

R^{67} = C_{1-6}alkyl, $(CH_2)_{0-6}$-C_{3-6}cycloalkyl, Ar-C_{0-6}alkyl, Het-C_{0-6}alkyl; R^{66} and R^{67} may be combined to form a 3-7 membered monocyclic or 7-10-membered bicyclic carbocyclic or heterocyclic ring, optionally substituted with 1-4 of C_{1-6}alkyl, Ph-C_{0-6}alkyl, Het-C_{0-6}alkyl, C_{1-6}alkoxy, Ph-C_{0-6}alkoxy, Het-C_{0-6}alkoxy, OH, $(CH_2)_{1-6}NR^{58}R^{59}$, $O(CH_2)_{1-6}NR^{58}R^{59}$;

R^{68} = R^{70} = R^{62}, R^{62}C(O), R^{62}C(S), R^{62}OC(O), R^{62}OC(O)NR^{59}CH(R^{71})(CO);

and pharmaceutically acceptable salts thereof.

The compounds of Formula I are hydrazidyl, bis-hydrazidyl and bis-aminomethyl carbonyl compounds having in common key structural features required of protease substrates, most particularly cathepsin K substrates. These structural features endow the present compounds with the appropriate molecular shape necessary to fit into the enzymatic active site, to bind to such active site,

thereby blocking the site and inhibiting enzymatic biological activity. Referring to Formula I, such structural features include the central electrophilic carbonyl, a peptidyl or peptidomimetic molecular backbone on either side of the central carbonyl, a terminal carbobenzyloxy moiety (e.g., Cbz-leucinyl), or a mimic thereof, on the backbone on one or both sides of the carbonyl, and optionally, an isobutyl side chain extending from the backbone on one or both sides of the carbonyl.

Abbreviations and symbols commonly used in the peptide and chemical arts are used herein to describe the compounds of the present invention. In general, the amino acid abbreviations follow the IUPAC-IUB Joint Commission on Biochemical Nomenclature as described in *Eur. J. Biochem.*, 158, 9 (1984). The term "amino acid" as used herein refers to the D- or L- isomers of alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine and valine.

"C_{1-6}alkyl" as applied herein is meant to include substituted and unsubstituted methyl, ethyl, n-propyl, isopropyl, n-butyl, isobutyl and t-butyl, pentyl, n-pentyl, isopentyl, neopentyl and hexyl and the simple aliphatic isomers thereof. Any C_{1-6}alkyl group may be optionally substituted independently by one or two halogens, SR', OR', $N(R')_2$, $C(O)N(R')_2$, carbamyl or C_{1-4}alkyl, where R' is C_{1-6}alkyl. C_0alkyl means that no alkyl group is present in the moiety. Thus, Ar-C_0alkyl is equivalent to Ar.

"C_{3-11}cycloalkyl" as applied herein is meant to include substituted and unsubstituted cyclopropane, cyclobutane, cyclopentane, cyclohexane, cycloheptane, cyclooctane, cyclononane, cyclodecane, cycloundecane.

"C_{2-6} alkenyl" as applied herein means an alkyl group of 2 to 6 carbons wherein a carbon-carbon single bond is replaced by a carbon-carbon double bond. C_{2-6}alkenyl includes ethylene, 1-propene, 2-propene, 1-butene, 2-butene, isobutene and the several isomeric pentenes and hexenes. Both cis and trans isomers are included.

"C_{2-6}alkynyl" means an alkyl group of 2 to 6 carbons wherein one carbon-carbon single bond is replaced by a carbon-carbon triple bond. C_{2-6} alkynyl includes acetylene, 1-propyne, 2-propyne, 1-butyne, 2-butyne, 3-butyne and the simple isomers of pentyne and hexyne.

"Halogen" means F, Cl, Br, and I.

"Ar" or "aryl" means phenyl or naphthyl, optionally substituted by one or more of Ph-C_{0-6}alkyl, Het-C_{0-6}alkyl, C_{1-6}alkoxy, Ph-C_{0-6}alkoxy, Het-C_{0-6}alkoxy, OH, $(CH_2)_{1-6}NR^{58}R^{59}$, $O(CH_2)_{1-6}NR^{58}R^{59}$; where R^{58}, R^{59} is H, C_{1-6}alkyl, Ar-C_{0-6}alkyl; Het-C_{0-6}alkyl, from C_{1-4}alkyl, OR', $N(R')_2$, SR', CF_3, NO_2, CN, CO_2R', CON(R'), F, Cl, Br and I.

As used herein "Het" or "heterocyclic" represents a stable 5- to 7-membered monocyclic or a stable 7- to 10-membered bicyclic heterocyclic ring, which is either saturated or unsaturated, and which consists of carbon atoms and from one to three heteroatoms selected from the group consisting of N, O and S, and wherein the nitrogen and sulfur heteroatoms may optionally be oxidized, and the nitrogen heteroatom may optionally be quaternized, and including any bicyclic group in which any of the above-defined heterocyclic rings is fused to a benzene ring. The heterocyclic ring may be attached at any heteroatom or carbon atom which results in the creation of a stable structure, and may optionally be substituted with one or two moieties selected from C_{1-4}alkyl, OR', $N(R')_2$, SR', CF_3, NO_2, CN, CO_2R', CON(R'), F, Cl, Br and I, where R' is C_{1-6}alkyl. Examples of such heterocycles include piperidinyl, piperazinyl, 2-oxopiperazinyl, 2-oxopiperidinyl, 2-oxopyrrolodinyl, 2-oxoazepinyl, azepinyl, pyrrolyl, 4-piperidonyl, pyrrolidinyl, pyrazolyl, pyrazolidinyl, imidazolyl, pyridyl, pyrazinyl, oxazolidinyl, oxazolinyl, oxazolyl, isoxazolyl, morpholinyl, thiazolidinyl, thiazolinyl, thiazolyl, quinuclidinyl, indolyl, quinolinyl, isoquinolinyl, benzimidazolyl, benzopyranyl, benzoxazolyl, furyl, pyranyl, tetrahydrofuryl, tetrahydropyranyl, thienyl, benzoxazolyl, thiamorpholinyl sulfoxide, thiamorpholinyl sulfone, and oxadiazolyl.

"HetAr" or "heteroaryl" means any heterocyclic moiety encompassed by the above definition of Het which is aromatic in character, e.g., pyridine.

It will be appreciated that the heterocyclic ring, , includes thiazoles, oxazoles, triazoles, thiadiazoles, oxadiazoles, isoxazoles, isothiazols, imidazoles, pyrazines, pyridazines, pyrimidines, triazines and tetrazines which are available by routine chemical synthesis and are stable. The single and double bonds (i.e., - -) in such heterocycles are arranged based upon the heteroatoms present so that the heterocycle is aromatic (e.g., it is a heteroaryl group). The term heteroatom as applied herein refers to oxygen, nitrogen and sulfur. When the heteroaryl group comprises a five membered ring, W is preferably an electron withdrawing group, such as halogen, -CN, $-CF_3$, $-NO_2$, $-COR^7$, $-CO_2R^6$, $-CONHR^6$, $-SO_2NHR^6$, -

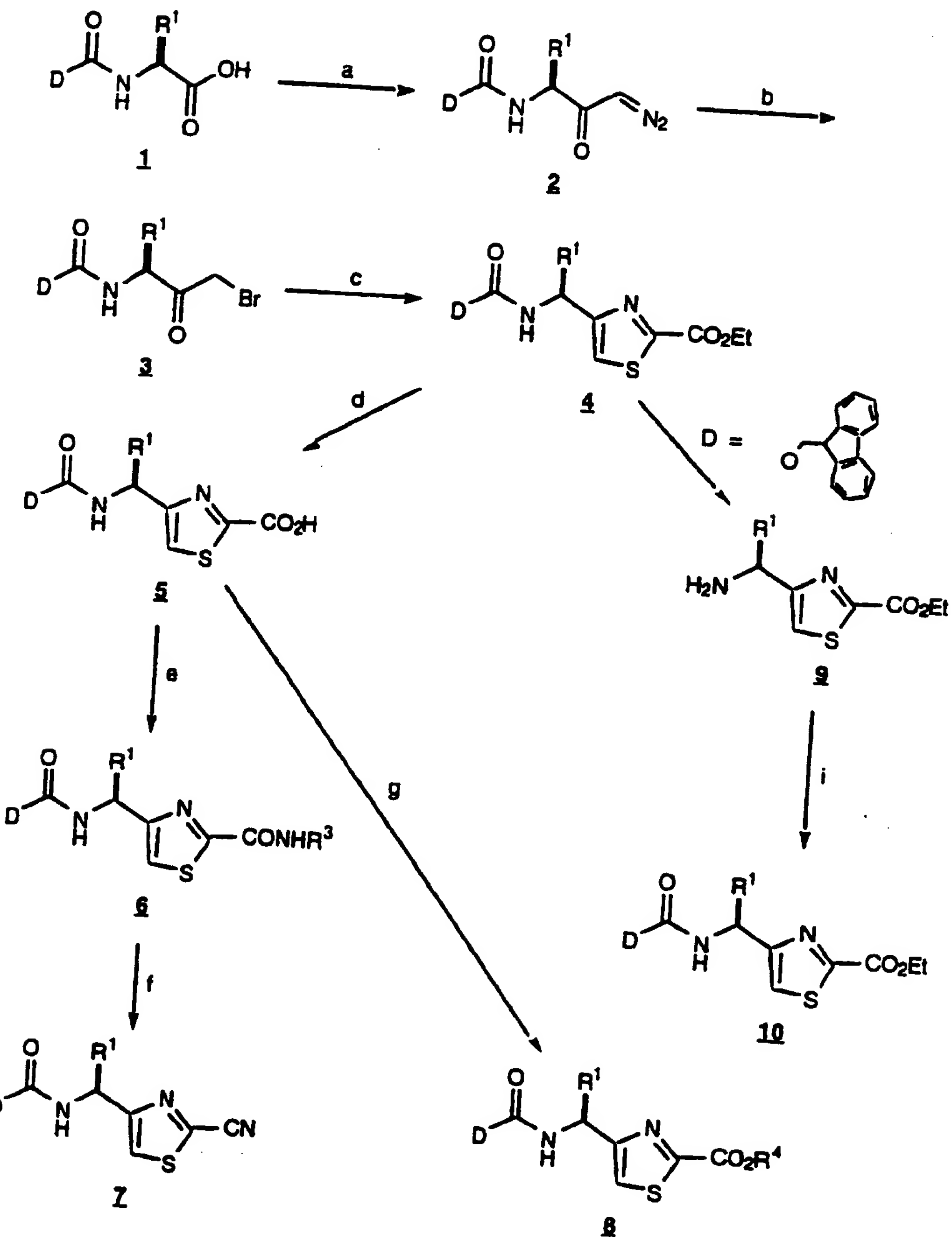
$NHSO_2R^6$, $-NHCOR^7$, $-O-COR^6$, $-SR^6$ or $NR'R^6$, or a similar electron withdrawing substituent as known in the art.

Certain radical groups are abbreviated herein. t-Bu refers to the tertiary butyl radical, Boc refers to the t-butyloxycarbonyl radical, Fmoc refers to the fluorenylmethoxycarbonyl radical, Ph refers to the phenyl radical, Cbz refers to the benzyloxycarbonyl radical.

Certain reagents are abbreviated herein. DCC refers to dicyclohexylcarbodiimide, DMAP is 2,6-dimethylaminopyridine, EDC refers to N-ethyl-N'(dimethylaminopropyl)-carbodiimide. HOBT refers to 1-hydroxybenzotriazole, DMF refers to dimethyl formamide, BOP refers to benzotriazol-1-yloxy-tris(dimethylamino)phosphonium hexafluorophosphate, DMAP is dimethylaminopyridine, Lawesson's reagent is 2,4-bis(4-methoxyphenyl)-1,3-dithia-2,4-diphosphetane-2,4-disulfide, NMM is N-methylmorpholine, TFA refers to trifluoroacetic acid, TFAA refers to trifluoroacetic anhydride and THF refers to tetrahydrofuran. Jones reagent is a solution of chromium trioxide, water, and sulfuric acid well-known in the art.

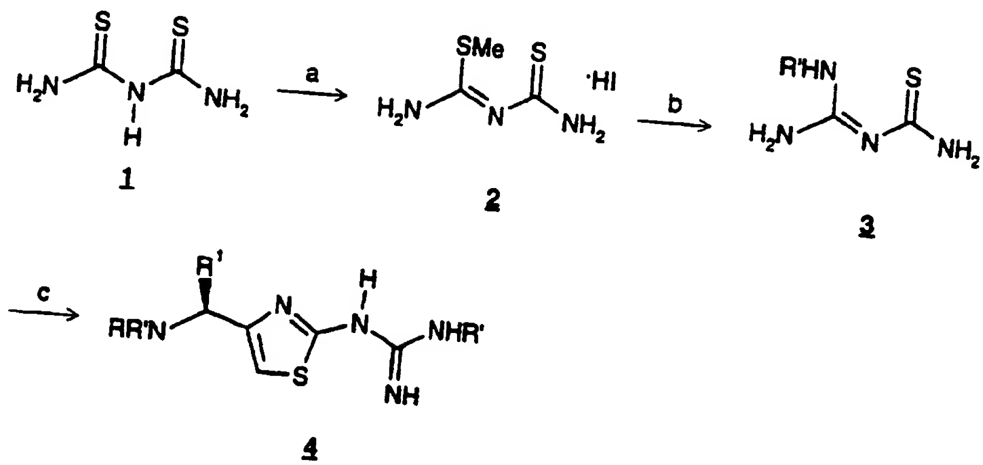
Compounds of formula (I) are prepared according to the methods detailed in Schemes 1-25.

Scheme 1



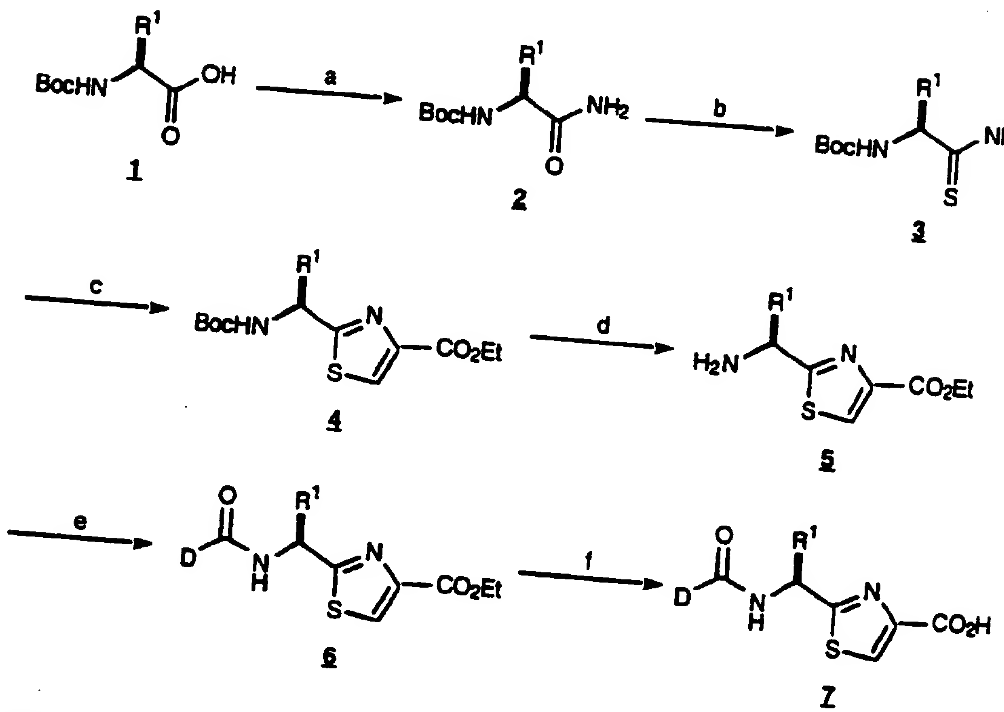
a) i-BuOCOCl, NMM, CH_2N_2, EtOAc, Et_2O; b) HBr, AcOH, EtOAc, Et_2O; c) H_2NCSCO_2Et, EtOH; d) NaOH, H_2O, THF; e) i-BuOCOCl, NMM, NH_2, THF or BOP, Et_3N, RNH_2, CH_2Cl_2; f) TFAA, pyridine, CH_2Cl_2; g) R^4OH, Boc_2O, Pyridine or R^4OH, EDCI, CH_2Cl_2; h) piperidine, DMF; i) BOP, Et_3N, D-CO_2H, CH_2Cl_2

Scheme 1A



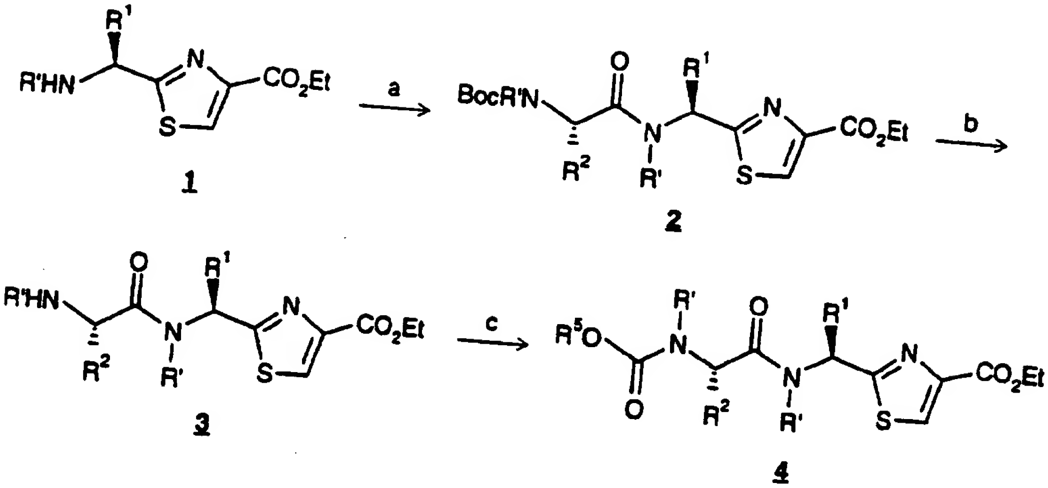
a) MeI, THF; b) $R'NH_2$, *i*-PrOH; c) Bromomethyl ketone, EtOH

Scheme 2



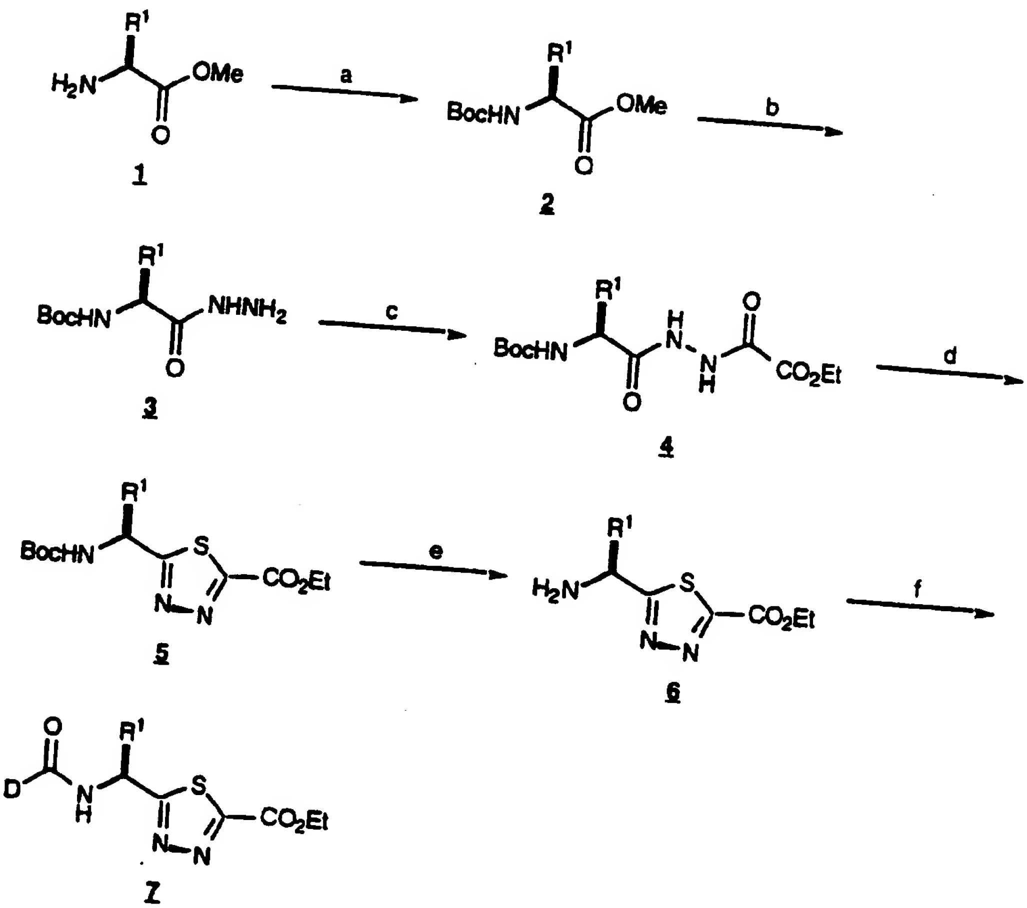
a) *i*-BuOCOCl, NMM, NH_3, THF; b) Lawesson's reagent, THF; c) $BrCH_2COCO_2Et$, TFAA, Pyridine, CH_2Cl_2; d) TFA; e) DCO_2H, EDC•HCl, HOBT, Et_3N, DMF; f) NaOH, H_2O, THF

Scheme 2A



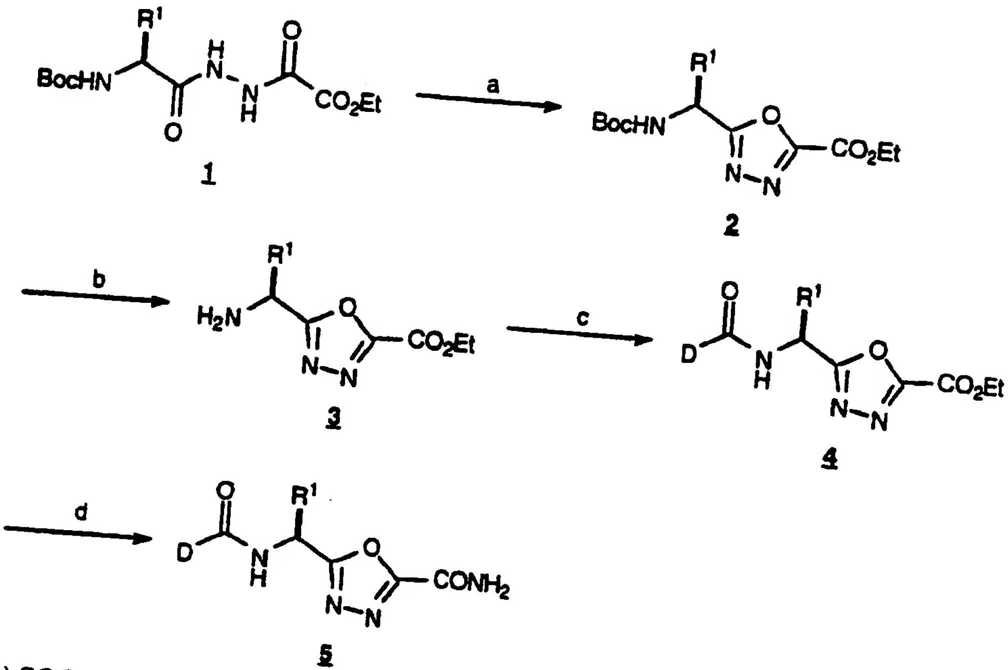
a) Boc-amino acid, EDC•HCl, 1-HOBT, DMF; b) TFA; c) R^5OCOCl, i-Pr_2NEt

Scheme 3



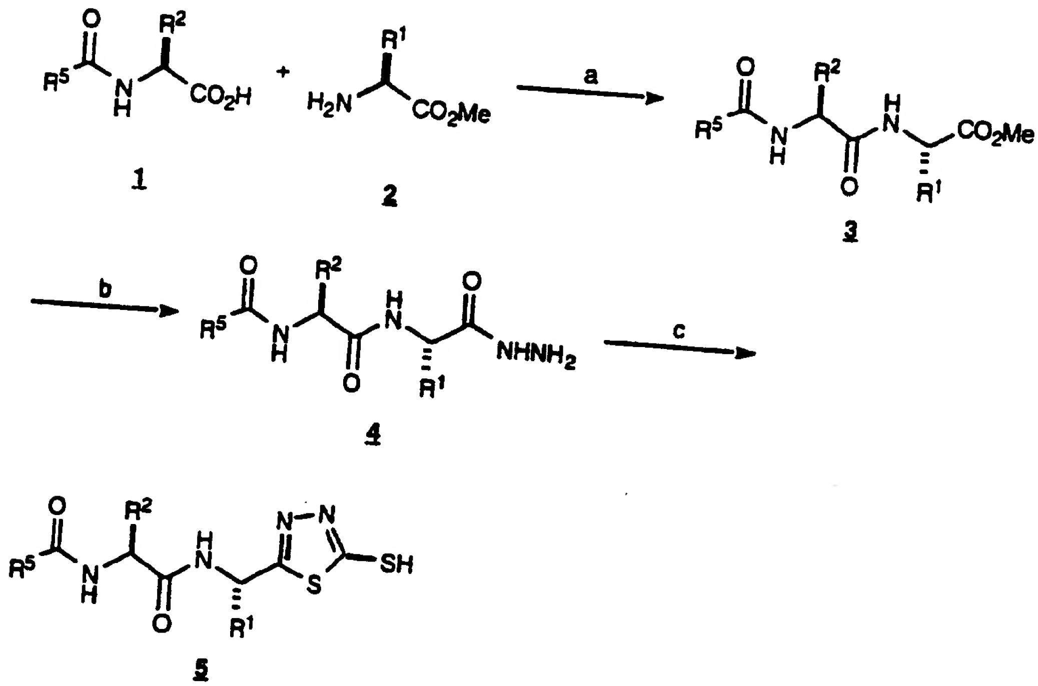
a) Boc_2O, Et_3N, THF; b) hydrazine hydrate, MeOH; c) EtO_2CCOCl, Pyridine, CH_2Cl_2; d) Lawesson's reagent, toluene; e) TFA, CH_2Cl_2; f) DCO_2H, EDC•HCl/HOBT, Et_3N, DMF

Scheme 4



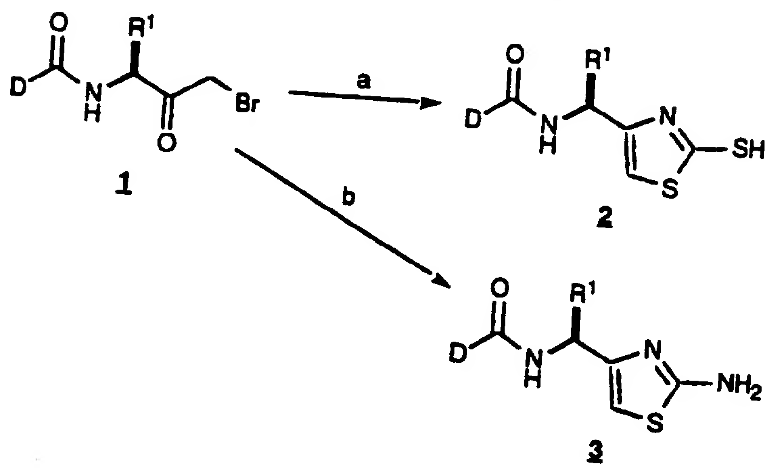
a) $SOCl_2$, pyridine, Et_2O, toluene; b) TFA, CH_2Cl_2; c) DCO_2H, EDC•HCl/HOBT, Et_3N, DMF; d) NH_3, EtOH

Scheme 5



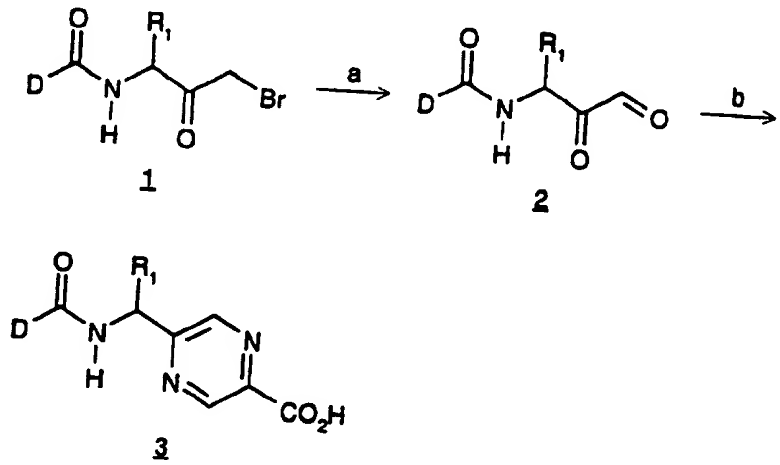
a) EDC•HCl/HOBT, Et_3N, DMF; b) H_2NNH_2•H_2O, MeOH; c) $CSCl_2$, Et_3N, $CHCl_3$

Scheme 6



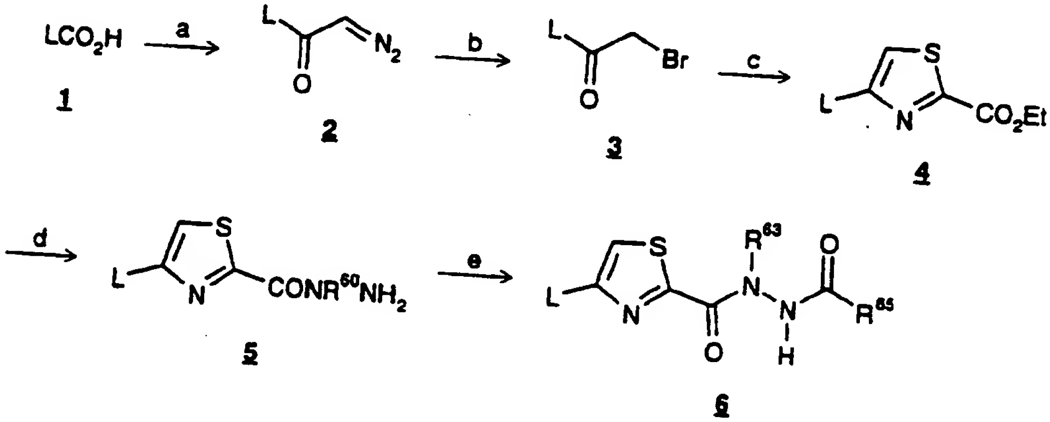
a) H_2NCS_2 NH_4^+, EtOH; b) H_2NCSNH_2, EtOH

Scheme 7



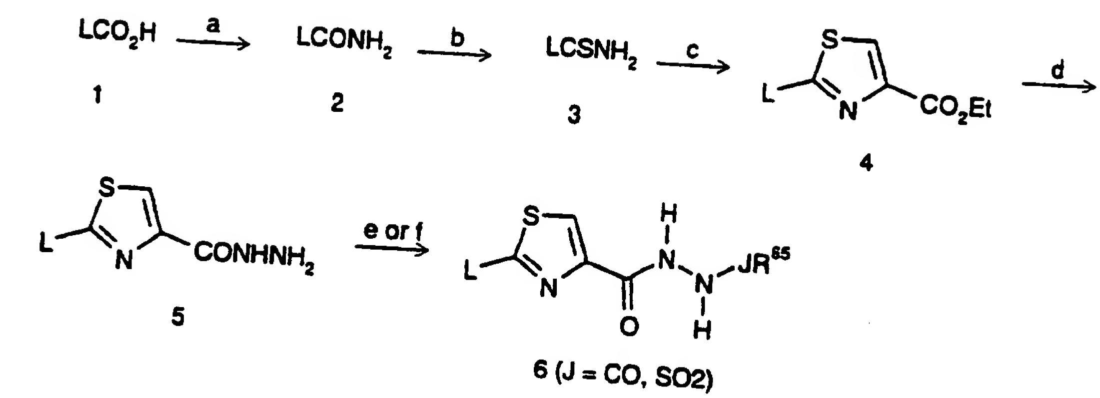
a) Et_2NO; b) $H_2NCH_2CH(NH_2)CO_2H$

Scheme 8



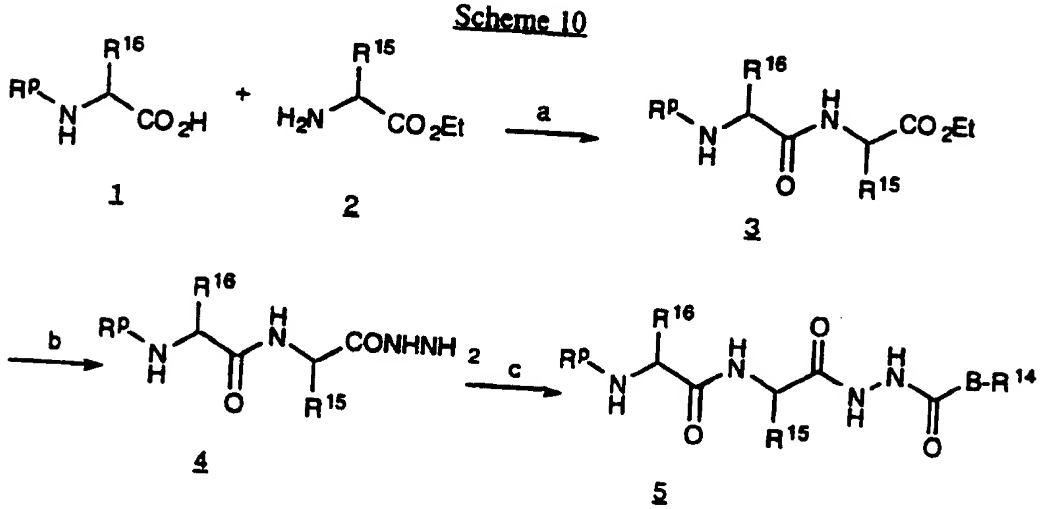
a) i. *i*-BuOCOCl, NMM, THF; ii. CH_2N_2, Et_2O; b) HBr, AcOH, Et_2O; c) H_2NCSCO_2Et, EtOH; d) $R^{63}NHNH_2$, EtOH; e) $R^{65}CO_2H$, EDC•HCl, 1-HOBT, DMF.

Scheme 9



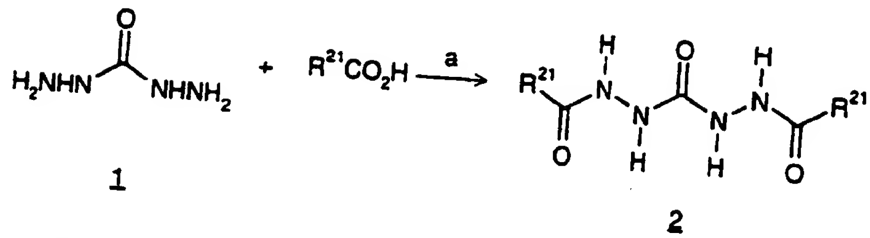
a) *i*-BuOCOCl, NMM, NH_3, THF; b) Lawesson's reagent, THF; c) i. EtO_2CCOCH_2Br; ii. TFAA, Py, CH_2Cl_2; d) $H_2NNH_2{\bullet}H_2O$, EtOH; e) $R^{65}SO_2Cl$, Py, CH_2Cl_2; f) $R^{65}CO_2H$, EDC•HCl, 1-HOBT, DMF.

Scheme 10



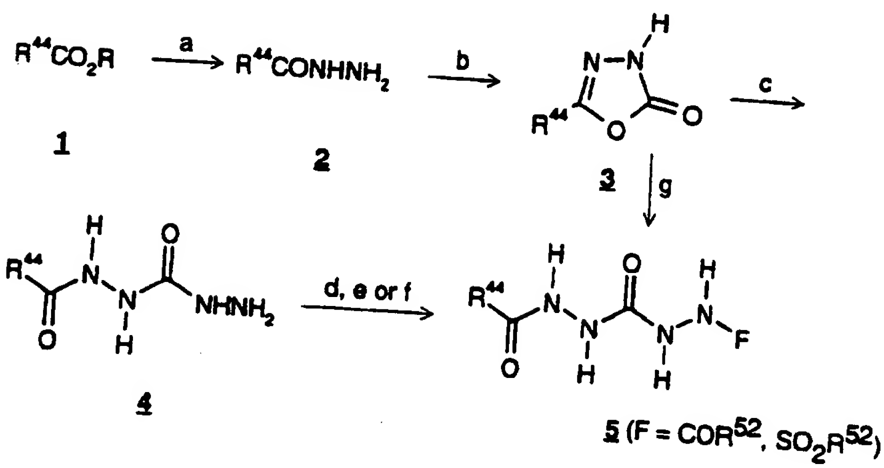
a) EDC•HCl, HOBT, DMF; b) $H_2NNH_2{\bullet}H_2O$, EtOH; c) R^{14}-B-CO_2H, EDC•HCL, HOBT, DMF

Scheme 11



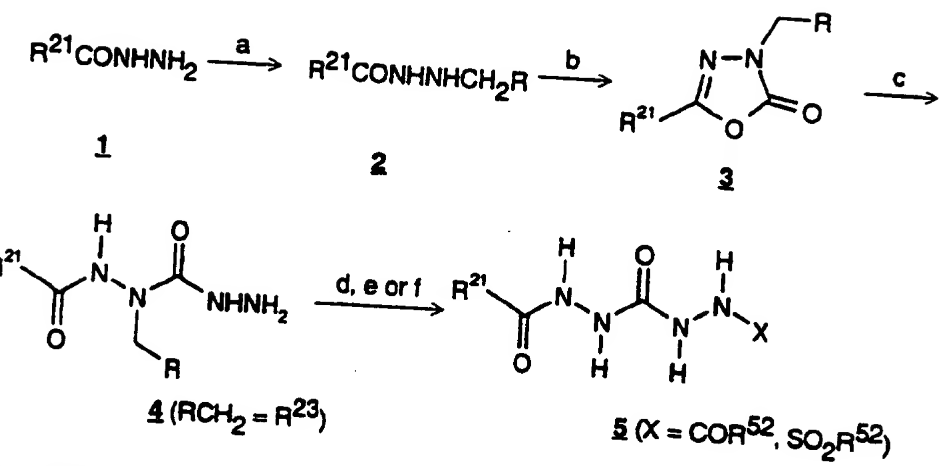
a) EDC.HCl, 1-HOBT, DMF

Scheme 12



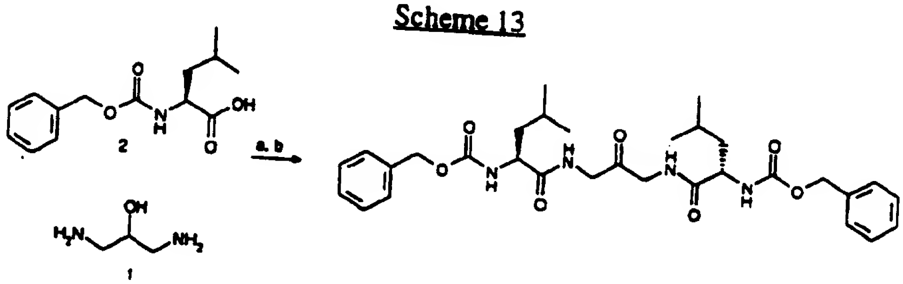
a) $H_2NNH_2 \cdot H_2O$, MeOH; b) Cl_2CO, PhMe; c) $H_2NNH_2 \cdot H_2O$, MeOH; d) $R^{49}CO_2H$, EDC·HCl, 1-HOBT, DMF; e) $R^{52}SO_2Cl$ or $R^{52}COCl$, pyridine, DMF; f) $R^{52}CO_2COR^{52}$; g) $R^{52}CONR^{51}NH_2$

Scheme 12A



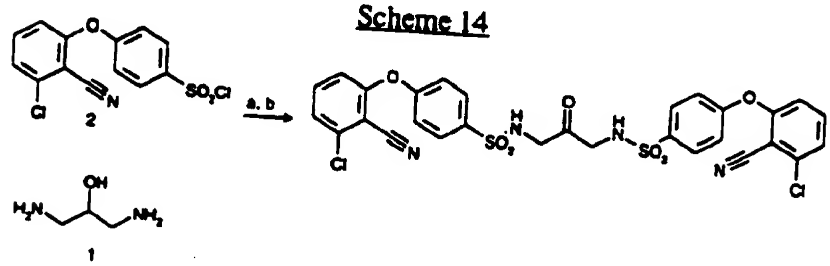
a) i. PhCHO, EtOH; ii. BH3·THF; b) Cl_2CO, PhMe; c) $H_2NNH_2 \cdot H_2O$, MeOH; d) $R^{52}CO_2H$, EDC·HCl, 1-HOBT, DMF; e) $R^{52}SO_2Cl$ or $R^{52}COCl$, pyridine, DMF; f) $R^{52}CO_2COR^{52}$

Scheme 13



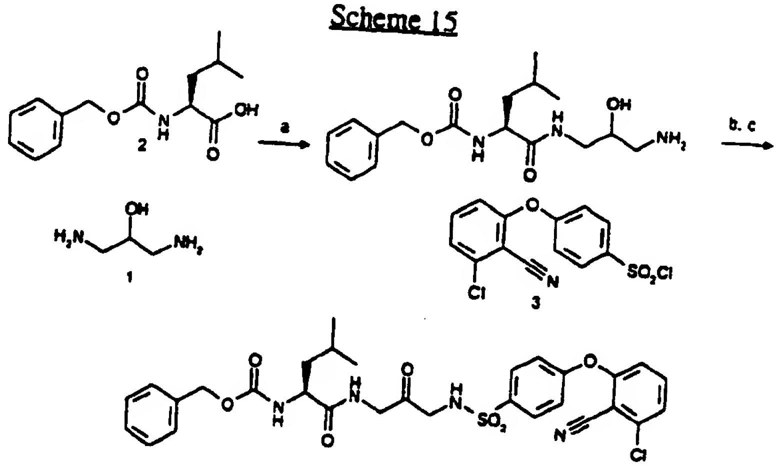
a) HBTU, NMM, DMF; b) Jones, acetone

Scheme 14



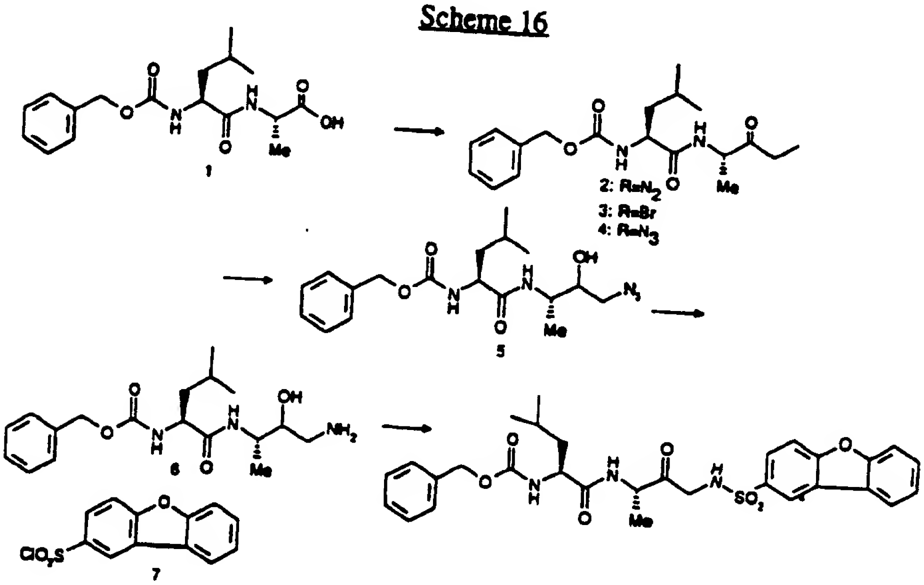
a) NMM, DMF; b) Jones, acetone

Scheme 15

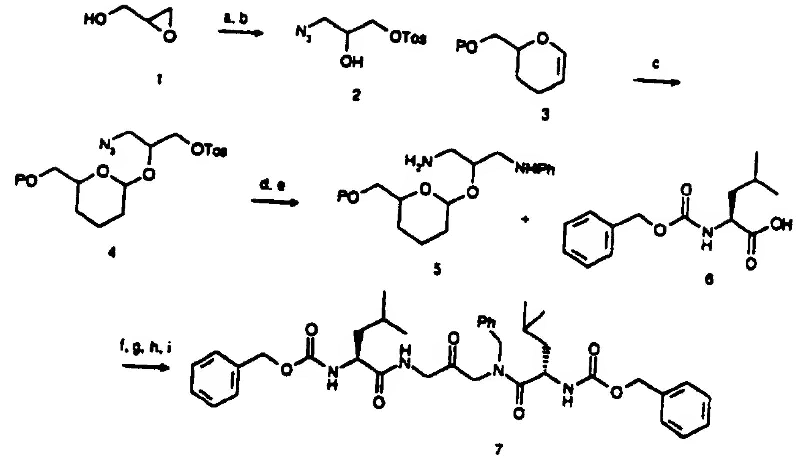


a) EDCI, HOBT, DMF; b) NMM, DMF, 3) Jones, acetone

Scheme 16

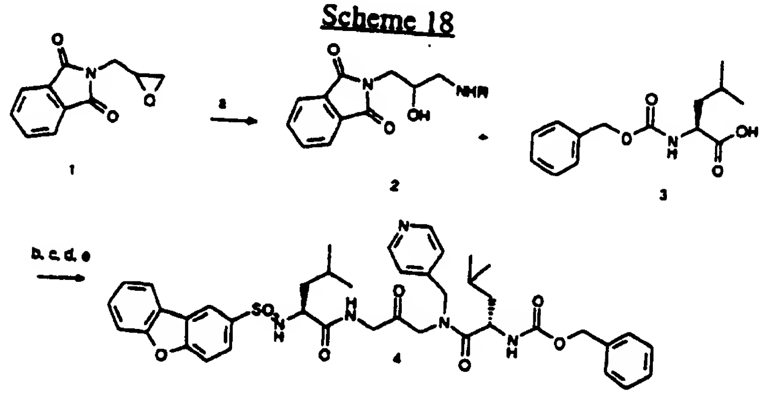


Scheme 17



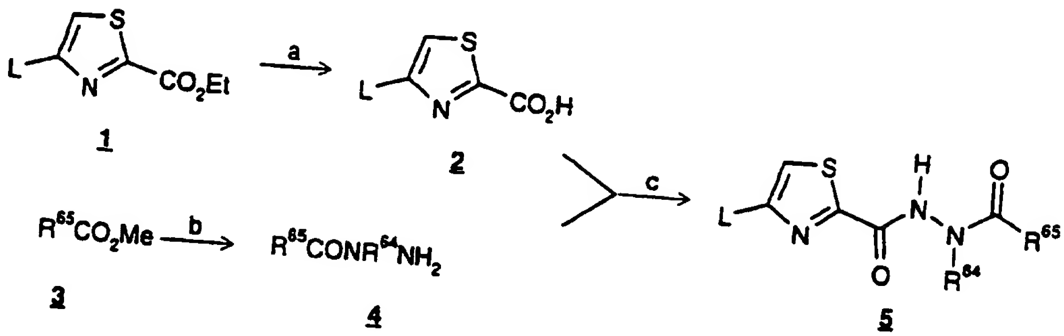
a) NaN_3, MeOH, H_2O; b) Tosyl chloride, triethylamine, CH_2Cl_2; c) Ellman dihydropyran resin (3), PPTS, $Cl(CH_2)_2Cl$; d) $PhCH_2NH_2$, toluene, 80 degrees C; e) HATU, N-methyl morpholine, NMP; f) $HS(CH_2)_3SH$, MeOH, Et_3N; g) Cbz-leucine (6), HBTU, N-methyl morpholine, NMP; h) TFA, CH_2Cl_2, Me_2S; i) Jones reagent, acetone

Scheme 18



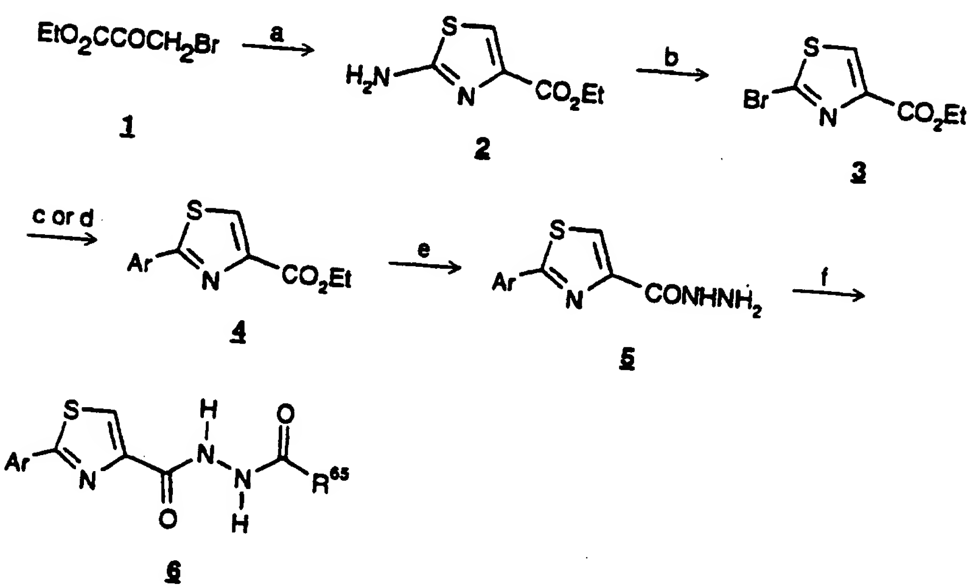
a) 4-pyridyl methyl amine, isopropanol, reflux; b) Cbz-leucine, HBTU, N-methyl morpholine, DMF; c) hydrazine, MeOH, reflux; d) 2-dibenzofuransulfonyl chloride, N-methyl morpholine, DMF; e) Jones reagent, acetone

Scheme 19



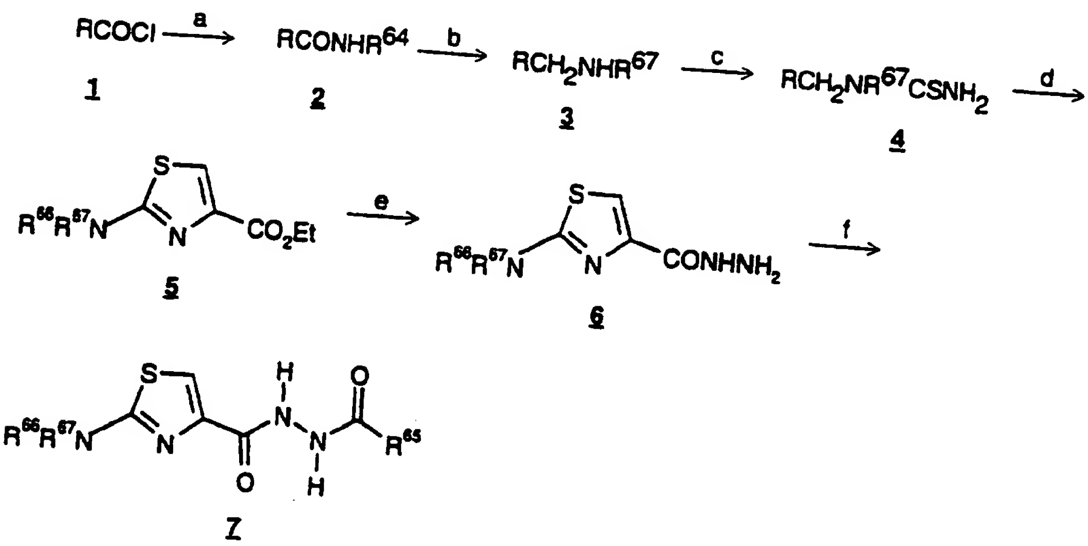
a) KOH, MeOH/H2O; b) $R^{66}NHNH_2$, EtOH; c) EDC•HCl, 1-HOBT, DMF

Scheme 20



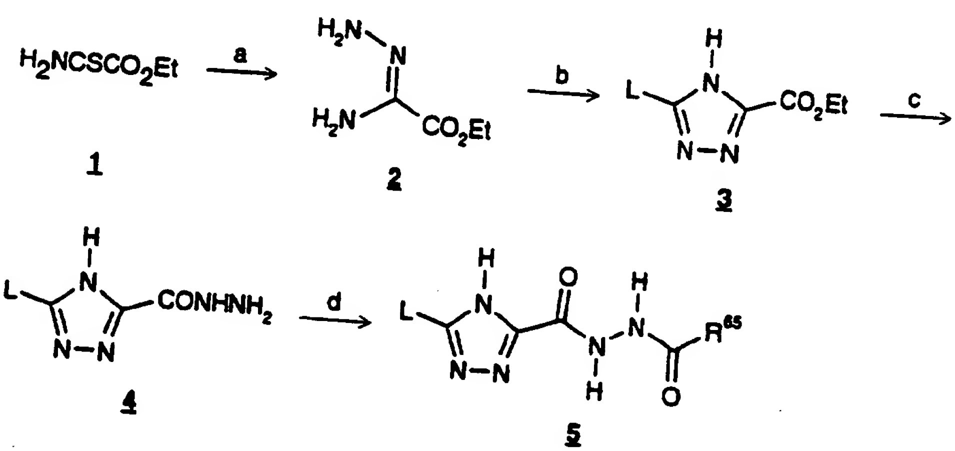
a) Thiourea, EtOH; b) i. $NaNO_2$, 16% aqueous HBr; ii. CuBr, 16% aqueous HBr; iii. HBr (cat.), EtOH; c) $ArB(OH)_2$, $Pd(PPh_3)_4$, CsF, DME; d) $ArSnMe_3$, $Pd(PPh_3)_4$, PhMe; e) $H_2NNH_2{\bullet}H_2O$, EtOH; e) $R^{65}CO_2H$, EDC•HCl, 1-HOBT, DMF.

Scheme 21



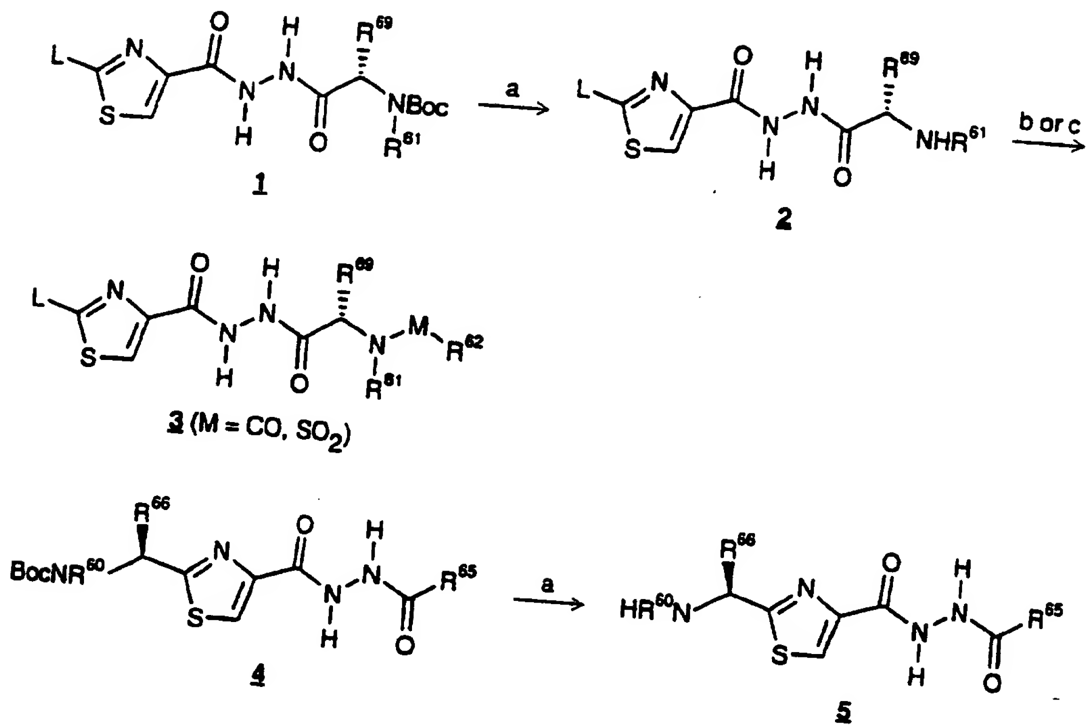
a) $R^{67}NH_2$, Py, CH_2Cl_2; b) $LiAlH_4$, THF; c) i. Cl_2CS, Py, CH_2Cl_2; ii. NH_3, MeOH or I. PhCONCS, $CHCl_3$; ii. K_2CO_3, MeOH, H_2O; d) EtO_2CCOCH_2Br, EtOH; e) H_2NNH_2•H_2O, EtOH; e) $R^{65}CO_2H$, EDC•HCl, 1-HOBT, DMF.

Scheme 22



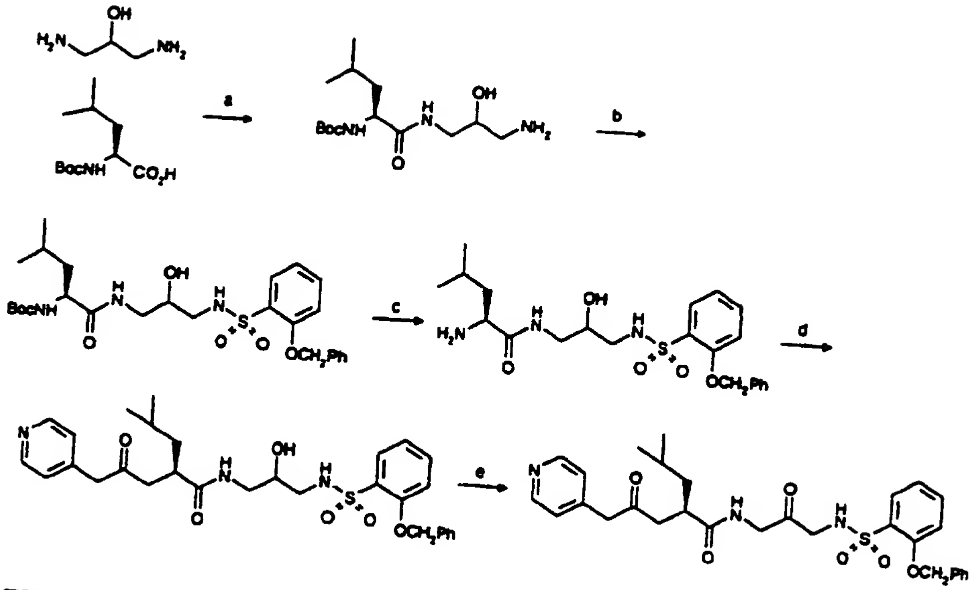
a) H_2NNH_2•H_2O, EtOH; b) LCO_2CO_2i-Bu, 200 °C; c) H_2NNH_2•H_2O, EtOH; d) $R^{65}CO_2H$, EDC•HCl, 1-HOBT, DMF

Scheme 23



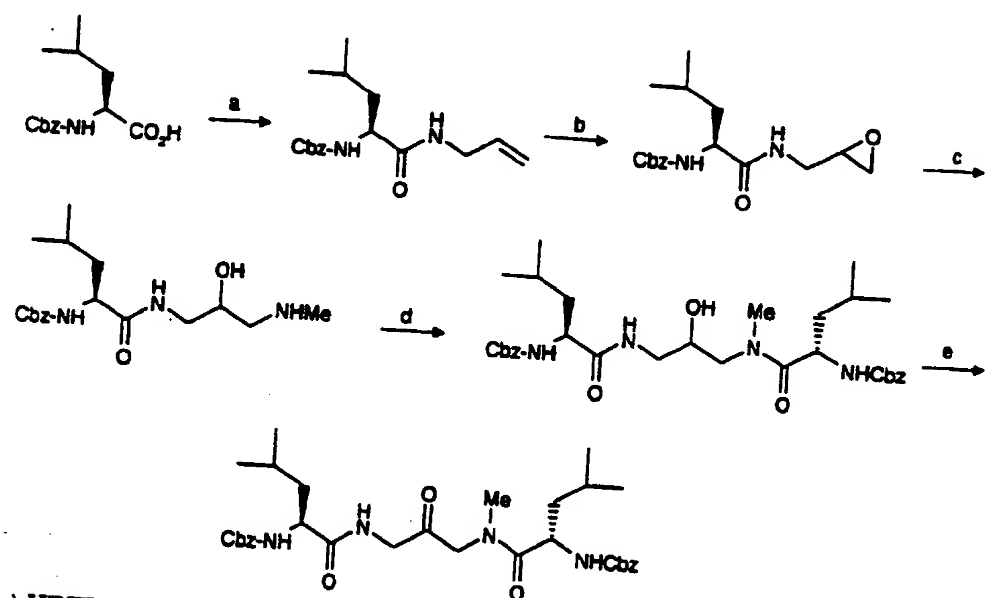
a) TFA; b) $R^{62}CO_2H$, EDC•HCl, 1-HOBT, DMF; c) $R^{62}SO_2Cl$, *i*-Pr_2NEt

Scheme 24



a) EDCI, DMF; b) 2-$PhCH_2OPhSO_2Cl$, NMM, DMF; c) TFA, DCM; d) 4-pyridyl acetic acid, HBTU, NMM, DMF; e) Jones

Scheme 25



a) HBTU, NMM, DMF; allyl amine; b) mCPBA, DCM; c) $MeNH_2$, isopropanol, 70 C; d) Cbz-leucine, EDCI, DMF; e) Jones, acetone

In another aspect, the present invention provides a novel cysteine protease in crystalline form, as defined by the positions in Table I herein.

In still another aspect, the present invention provides a novel protease composition characterized by a three dimensional catalytic site formed by the atoms of the amino acid residues listed in Table XXIX herein.

The three dimensional (3D) structure of the instant protease reveals that human cathepsin K is highly homologous to other known cysteine proteinases of the papain family. Cathepsin-K folds into two subdomains separated by the active site cleft, a characteristic of the papain family of cysteine proteases. The overall fold of cathepsin K is very similar to that of papain and actinidin. There is an insertion of one additional residue in cathepsin K at residue alanine 79 compared to papain. This insertion is easily accommodated in the turn at the carboxy terminal end of the helix formed by residues methionine 68-lysine 77 of cathepsin K. There is a different conformation for the backbone atoms of residues asparagine 99 to lysine 103 at the surface of cathepsin K compared to that in papain. Other differences in the backbone conformations between cathepsin K and papain are: a two residue insertion in loop residues 126-127, a two residue insertion at residue aspartate 152, the insertion of 4 residues at glutamine 172 and a difference in the conformation of the loop around residue lysine 200. There are many more differences in the structure of human cathepsin K and human cathepsin B, however, the secondary structure is preserved well between these two enzymes.

Listed in Figure 1 are the known amino acid sequences for the papain superfamily of cysteine proteases cathepsin K, cathepsin S, cathepsin L, papain, actinidin, cathepsin H and cathepsin B, aligned to illustrate the homologies there between.

According to the present invention the crystal structure of human cathepsin K has been determined in the absence of inhibitor and in complex with nine separate inhibitors at resolutions from 3.0 to 2.2 Ångstroms. The structures were determined using the method of molecular replacement and refined to R_c values ranging from 0.190-0.267 with the exception of the enzyme in the absence of inhibitor which was not refined.

Further refinement of the atomic coordinates will change the numbers in Table I. Refinement of the crystal structure from another crystal form will result in a new set of coordinates, determination of the crystal structure of another cysteine

protease will also result in different set of numbers for coordinates in Table I which has an experimental error of approximately 0.4 Ångstroms. Also for example, the amino acid sequence of the cysteine proteases can be varied by mutation derivatization or by use of a different source of the protein.

Human cathepsin K contains 215 amino acids and the model of the enzyme provided herein is represented by all 215 residues.

The cathepsin K crystal structure reveals an active site that is heretofor unknown and comprises a distinct three dimensional arrangement of atoms.

Table I discloses the protein coordinates of cathepsin K. These data are reported for the crystal structures described herein. The data are reported in Ångstroms with reference to an orthogonal coordinate system in standard format, illustrating the atom, i.e., nitrogen, oxygen, carbon, sulfur (at α, ß, γ, δ, or ε, positions in the amino acid residues); the amino acid residue in which the atom is located with amino acid number, and the coordinates X, Y and Z in Angstroms (Å) from the crystal structure. Note that each atom in the active site and the entire structure has an unique position in the crystal. The data also report the B or Temperature Factor values, which indicate the degree of thermal motion of the atom in root mean square displacement measurements ($Å^2$). Figure 2 illustrates the cathepsin K structure of the invention, including the active site.

The active site of cathepsin K bound to E-64 is shown in Figure 3. The conformation of E-64 bound to cathepsin K resembles that seen in the published structures of the papain-E-64 complex (Varughese, K.I., *Biochemistry* **28**, 1330-1332 (1989)) and actinidin-E-64 Varughese, K.I. , *Biochemistry* **31**, 5172-5176 (1992)). The covalent bond between the sulfur of cysteine 25 and the carbon C2 of the inhibitor is very clear in the electron density. Differences in the sidechain atoms lining the active site pockets on the enzyme of the various members of the papain family of cysteine proteases give rise to different interactions between the atoms of E-64 and the protein in these structures. In cathepsin K, the isobutyl atoms of the leucine lie well buried in the hydrophobic pocket formed by the side chain atoms of the cathepsin K residues leucine 160, alanine 134 and methionine 68 shielding these atoms of E-64 from solvent. In papain the leucyl side chain atoms of E-64 do not penetrate as deeply into this hydrophobic pocket. Another pocket of cathepsin K is occupied by the guanidinium atoms of E-64. A hydrogen bond forms between N4 of E-64 and the backbone carbonyl oxygen of glutamate 59 and the OD2 oxygen of aspartate 61. The carboxylate oxygen of aspartate 61 also makes a hydrogen bond

with the N3 atom of E-64. The sidechain atoms of aspartate 61 lie at the entrance to this pocket in cathepsin K. These interactions are not possible in papain because the corresponding residue in papain is tyrosine 61 which blocks access. The carboxylate oxygens of E-64 make hydrogen bonding interactions with the ND1 atom of histidine 162 and the NE2 atom of glutamine 19. These interactions are also seen in papain and actinidin. The atoms of E-64 do not penetrate the complete region of the enzyme active site. As in papain, the backbone nitrogen atoms of residue glycine 66 in cathepsin K makes a hydrogen bond with the carbonyl oxygen atom O4 of the E-64. Also, the carbonyl oxygen of glycine 66 of cathepsin K forms a hydrogen bond with N2 of E-64. A portion of the regions of the active site are very similar in conformation in cathepsin K, papain and actinindin. A comparison of the active site of cathepsin K and cathepsin B reveals many more differences than observed in comparing papain or actinidin to cathepsin K. A portion of the active site of cathepsin B differs significantly from the corresponding portion of the active site in cathepsin K. The presence of the loop glutamate 107 - proline 116 in human cathepsin B is presumed responsible for the dipeptidyl carboxypeptidase activity of this enzyme and has no equivalent in cathepsin K, papain or actinidin. This loop makes this region of the active site of cathepsin B much smaller than in the other members of this papain family of cysteine proteases including cathepsin K. Despite the differences between the active sites of human cathepsin B and cathepsin K, the active site cysteine residues are almost exactly superimposed by an alignment of structurally homologous alpha carbon atoms in cathepsin B and cathepsin K. Differences in the hydrophobic pocket near leucine 160 in cathepsin K are also evident in cathepsin B. The residues forming this pocket are replaced by proline 78 in place of methionine 68 in cathepsin K and glutamate 243 in cathepsin B is structurally equivalent to leucine 160 in cathepsin K. Interestingly, the residues whose sidechain atoms form hydrogen bonds to the E-64 inhibitor in cathepsin K, namely histidine 162, glutamine 19 and aspartate 61, have structurally homologous residues in cathepsin B, namely histidine 197, glutamine 23 and aspartate 67 respectively.

Specific interactions of certain inhibitors of the present invention at the active site of cathepsin K are detailed hereinbelow.

3 (S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone makes hydrophobic contacts with the enzyme residues indole ring of tryptophan 184 and the sidechain atom CG of glutamine 19. Oxygen O26 forms a

bifurcated hydrogen bond with the amide nitrogen of cysteine 25 and the NE2 atom of glutamine 19. The active site nucleophilic sulfur of residue cysteine 25 is covalently linked to carbon C25 of the inhibitor, which adopts a tetrahedral conformation.

Bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one exhibits the same interaction as 3 (S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone; carbon C21 of this inhibitor is covalently linked to SG of cysteine 25. The isopropyl atoms CC34,C35,C36 and C37 of the inhibitor form hydrophobic interactions with the sidechain atoms of residues on the enzyme surface, which form a hydrophobic pocket. This pocket is formed by atoms from methionine 68, leucine 209, alanine 163 and alanine 134 and portions of tyrosine 67.

2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide has interactions similar to bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one and, in addition, the atoms C23-29 of the inhibitor CBZ group make an edge-face stacking interaction with the phenol ring of tyrosine 67. Inhibitor atom C21 is covalently bound the enzyme.

The sulfur atom of (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide contacts the ND1 atom of histidine 163 and the indole ring of tryptophan 184. Carbon C22 is covalently attached to SG of cysteine 25.

The CBZ atoms C20-26 of 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide interact with the sidechain atoms of leucine 160. Carbon C19 is covalently attached to SG of cysteine 25.

Cathepsin K binds selectively one stereoisomer of 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone. Carbon C22 is covalently attached to SG of cysteine 25. Atoms C14 and C15 of the inhibitor 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone form hydrophobic contacts with the sidechain atoms of glutamine 143 and asparagine 161 and the mainchain of alanine 137 and serine 138.

4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone interacts in a similar manner to 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone. Again one stereoisomer is bound. Carbon C17 is covalently attached to SG of cysteine 25. The interaction of 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone is

the same as for 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone, except carbon C22 is covalently attached to SG of cysteine 25.

Atom O24 of 1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one forms a hydrogen bond interaction with the amide NH of glycine 66. Carbon C19 is covalently attached to SG of cysteine 25.

In summary, all inhibitors exhibit an aromatic interaction with atoms of the indole of Tryptophan 184. Isopropyl atoms C12-15 of 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide and (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide make hydrophobic contacts with main chain atoms of residues glutamine 21, cysteine 22 and glycine 23. The NE2 atom of glutamine 19 is able to donate a hydrogen bond to oxygen atom 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide:O22, 1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one:O20, 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide:O20, 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone:O23, bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one:O22, 3(S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone:O26, 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone:O42, (1S, 2'R)-N-2-[[(1-benzyloxycarbonyl)amino]-3-methylbutyl]thiazol-4-ylcarbonyl-N'-2'-(benzyloxycarbonyl)amino-4'-methylpenanoylhydrazide:O23, 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone:O23. The backbone amide nitrogen of glycine 66 donates a hydrogen bond to 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide:O39, 1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one:O24, 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide:O37, 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone:O40, bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one:O39, (1S, 2'R)-N-2-[[(1-benzyloxycarbonyl)amino]-3-methylbutyl]thiazol-4-ylcarbonyl-N'-2'-(benzyloxycarbonyl)amino-4'-methylpenanoylhydrazide:O40, 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone:O31. The hydrophobic pocket lined with atoms from residues methionine 68, leucine 209, alanine 163 and alanine 134 and portions

of tyrosine 67 interact with the isopropyl atoms; bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one:C34-37, 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide: C34-37, (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide; :C35-38, 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide:C32-35, 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone:C35-38, 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone:C19-22, 1 -N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one:C26-29. All inhibitors except 3(S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone and 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone have aromatic groups that interact with tyrosine 67 on the protein. All inhibitors are covalently linked to the cysteine 25 SG atom through an inhibitor carbon atom.

The crystal structure of the protease of the present invention reveals the three dimensional structure of novel active site formed by the atoms of the amino acid residues listed in Table XXIX.

This structure is clearly useful in the structure-based design of protease inhibitors, which may be used as therapeutic agents against diseases in which inhibition of bone resorption is indicated. The discovery of the novel cathepsin K catalytic site permits the design of potent, highly selective protease inhibitors.

Another aspect of this invention involves a method for identifying inhibitors of cathepsin K characterized by the crystal structure and novel active site described herein, and the inhibitors themselves. The novel protease crystal structure of the invention permits the identification of inhibitors of protease activity. Such inhibitors may bind to all or a portion of the active site of cathepsin K; or even be competitive or non-competitive inhibitors. Once identified and screened for biological activity, these inhibitors may be used therapeutically or prophylactically to block protease activity.

One design approach is to probe the cathepsin K of the invention with molecules composed of a variety of different chemical entities to determine optimal sites for interaction between candidate cathepsin K inhibitors and the enzyme. For example, high resolution X-ray diffraction data collected from crystals saturated with solvent allows the determination of where each type of solvent molecule sticks.

Small molecules that bind tightly to those sites can then be designed and synthesized and tested for their cathepsin K inhibitor activity.

This invention also enables the development of compounds that can isomerize to short-lived reaction intermediates in the chemical reaction of a substrate or other compound that binds to or with cathepsin K. Thus, the time-dependent analysis of structural changes in cathepsin K during its interaction with other molecules is permitted. The reaction intermediates of cathepsin K can also be deduced from the reaction product in co-complex with cathepsin K. Such information is useful to design improved analogues of known cysteine protease inhibitors or to design novel classes of inhibitors based on the reaction intermediates of the cathepsin K enzyme and cathepsin K inhibitor co-complex. This provides a novel route for designing cathepsin K inhibitors with both high specificity and stability.

Another approach made possible by this invention, is to screen computationally small molecule data bases for chemical entities or compounds that can bind in whole, or in part, to the cathepsin K enzyme. In this screening, the quality of fit of such entities or compounds to the binding site may be judged either by shape complementarity [R. L. DesJarlais et al., J. Med. Chem. 31:722-729 (1988)] or by estimated interaction energy [E. C. Meng et al, J. Comp. Chem., 13:505-524 (1992)].

Because cathepsin K may crystallize in more than one crystal form, the structure coordinates of cathepsin K, or portions thereof, as provided by this invention are particularly useful to solve the structure of those other crystal forms of cathepsin K. They may also be used to solve the structure of cathepsin K mutants, cathepsin K co-complexes, or of the crystalline form of any other protein with significant amino acid sequence homology to any functional domain of cathepsin K.

One method that may be employed for this purpose is molecular replacement. In this method, the unknown crystal structure, whether it is another crystal form of cathepsin K, a cathepsin K mutant, or a cathepsin K co-complex, or the crystal of some other protein with significant amino acid sequence homology to any functional domain of cathepsin K, may be determined using the cathepsin K structure coordinates of this invention as provided in Table I. This method will provide an accurate structural form for the unknown crystal more quickly and efficiently than attempting to determine such information *ab initio*.

Thus, the cathepsin K structure provided herein permits the screening of known molecules and/or the designing of new molecules which bind to the protease structure, particularly at the active site, via the use of computerized evaluation systems. For example, computer modeling systems are available in which the sequence of the protease, and the protease structure (i.e., atomic coordinates of cathepsin K and/or the atomic coordinate of the active site cavity, bond angles, dihedral angles, distances between atoms in the active site region, etc. as provided by Table I may be input. Thus, a machine readable medium may be encoded with data representing the coordinates of Table I in this process. The computer then generates structural details of the site into which a test compound should bind, thereby enabling the determination of the complementary structural details of said test compound.

More particularly, the design of compounds that bind to or inhibit cathepsin K according to this invention generally involves consideration of two factors. First, the compound must be capable of physically and structurally associating with cathepsin K. Non-covalent molecular interactions important in the association of cathepsin K with its substrate include hydrogen bonding, van der Waals and hydrophobic interactions.

Second, the compound must be able to assume a conformation that allows it to associate with cathepsin K. Although certain portions of the compound will not directly participate in this association with cathepsin K, those portions may still influence the overall conformation of the molecule. This, in turn, may have a significant impact on potency. Such conformational requirements include the overall three-dimensional structure and orientation of the chemical entity or compound in relation to all or a portion of the binding site, e.g., active site or accessory binding site of cathepsin K, or the spacing between functional groups of a compound comprising several chemical entities that directly interact with cathepsin K.

The potential inhibitory or binding effect of a chemical compound with cathepsin K may be estimated prior to its actual synthesis and testing by the use of computer modeling techniques. If the theoretical structure of the given compound suggests insufficient interaction and association between it and cathepsin K, synthesis and testing of the compound is obviated. However, if computer modeling indicates a strong interaction, the molecule may then be synthesized and tested for

its ability to bind to cathepsin K in a suitable assay. In this manner, synthesis of inoperative compounds may be avoided.

An inhibitory or other binding compound of cathepsin K may be computationally evaluated and designed by means of a series of steps in which chemical entities or fragments are screened and selected for their ability to associate with the individual binding pockets or other areas of cathepsin K.

One skilled in the art may use one of several methods to screen chemical entities or fragments for their ability to associate with cathepsin K and more particularly with the individual binding pockets of the cathepsin K active site or accessory binding site. This process may begin by visual inspection of, for example, the active site on the computer screen based on the cathepsin K coordinates in Table I. Selected fragments or chemical entities may then be position cathepsin K. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics forcefields, such as CHARMM and AMBER.

Specialized computer programs may also assist in the process of selecting fragments or chemical entities. These include:

- GRID [P. J. Goodford, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem., 28:849-857 (1985)]. GRID is available from Oxford University, Oxford, UK.

- MCSS [A. Miranker and M. Karplus, "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method", Proteins: Structure, Function and Genetics, 11:29-34 (1991)]. MCSS is available from Molecular Simulations, Burlington, MA.

- AUTODOCK [D. S. Goodsell and A. J. Olsen, "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8:195-202 (1990)]. AUTODOCK is available from Scripps Research Institute, La Jolla, CA.

- DOCK [I. D. Kuntz et al, "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161:269-288 (1982)]. DOCK is available from University of California, San Francisco, CA.

Additional commercially available computer databases for small molecular compounds includes Cambridge Structural Database and Fine Chemical Database, for a review see Rusinko, A., Chem. Des. Auto. News 8, 44-47 (1993).

Once suitable chemical entities or fragments have been selected, they can be assembled into a single compound or inhibitor. Assembly may be proceeded by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of cathepsin K. This would be followed by manual model building using software such as Quanta or Sybyl.

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include:

- CAVEAT [P. A. Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc. 78, pp. 182-196 (1989)]. CAVEAT is available from the University of California, Berkeley, CA.
- 3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, CA). This area is reviewed in Y. C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35:2145-2154 (1992).
- HOOK (available from Molecular Simulations, Burlington, MA).

Instead of proceeding to build a cathepsin K inhibitor in a step-wise fashion one fragment or chemical entity at a time as described above, inhibitory or other type of binding compounds may be designed as a whole or "*de novo*" using either an empty active site or optionally including some portion(s) of a known inhibitor(s). These methods include:

- LUDI [H.-J. Bohm, "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", J. Comp. Aid. Molec. Design, 6:61-78 (1992)]. LUDI is available from Biosym Technologies, San Diego, CA.
- LEGEND [Y. Nishibata and A. Itai, Tetrahedron, 47:8985 (1991)]. LEGEND is available from Molecular Simulations, Burlington, MA.
- LeapFrog (available from Tripos Associates, St. Louis, MO).

Other molecular modeling techniques may also be employed in accordance with this invention. See, e.g., N. C. Cohen et al, "Molecular Modeling Software and Methods for Medicinal Chemistry", J. Med. Chem., 33:883-894 (1990). See also, M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2:202-210 (1992). For example, where the structures of test compounds are known, a model of the test compound may be superimposed over the model of the structure of the invention. Numerous

methods and techniques are known in the art for performing this step, any of which may be used. See, e.g., P.S. Farmer, Drug Design, Ariens, E.J., ed., Vol. 10, pp 119-143 (Academic Press, New York, 1980); U.S. Patent No. 5,331,573; U.S. Patent No. 5,500,807; C. Verlinde, Structure, 2:577-587 (1994); and I. D. Kuntz, Science, 257:1078-1082 (1992). The model building techniques and computer evaluation systems described herein are not a limitation on the present invention.

Thus, using these computer evaluation systems, a large number of compounds may be quickly and easily examined and expensive and lengthy biochemical testing avoided. Moreover, the need for actual synthesis of many compounds is effectively eliminated.

Once identified by the modeling techniques, the protease inhibitor may be tested for bioactivity using standard techniques. For example, structure of the invention may be used in binding assays using conventional formats to screen inhibitors. Suitable assays for use herein include, but are not limited to, the enzyme-linked immunosorbent assay (ELISA), or a fluoresence quench assay. See, for example, the cathepsin K activity assay of Example 2 below. Other assay formats may be used; these assay formats are not a limitation on the present invention.

In another aspect, the protease structure of the invention permit the design and identification of synthetic compounds and/or other molecules which have a shape complimentary to the conformation of the protease active site of the invention. Using known computer systems, the coordinates of the protease structure of the invention may be provided in machine readable form, the test compounds designed and/or screened and their conformations superimposed on the structure of the protease of the invention. Subsequently, suitable candidates identified as above may be screened for the desired protease inhibitory bioactivity, stability, and the like.

Once identified and screened for biological activity, these inhibitors may be used therapeutically or prophylactically to block cathepsin K activity.

The following examples illustrate various aspects of this invention. These examples do not limit the scope of this invention which is defined by the appended claims.

EXAMPLE 1: Analysis of the Structure of Cathepsin K

A. *Expression, Purification and Crystallization*

Cathepsin K (see Fig. 1) was expressed and purified as described in Bossard, M. J., et al., *J. Biol. Chem.* **271**, 12517-12524 (1996).

Crystals of cathepsin K were grown by vapor diffusion in hanging drops from a solution of 30% PEG 8000, 0.1 M Na^+/K^+ phosphate at pH 4.5 containing 0.2M Li_2SO_4. Crystals of the complex are tetragonal, space group $P4_32_12$, with cell constants of a=57.7 Ångstroms and c=131.1 Ångstroms. The crystals contain one molecule in the asymmetric unit and contain 36 % solvent with a V_m value of 2.3 $Å^3$/Dalton. The structure was determined by molecular replacement using X-PLOR [Brunger, A.T., et al., *Science*, **235**, 458-460 (1987)]. The starting model consisted of the protein atoms from the cathepsin K E-64 complex structure described herein.

B. ***Model Building and Refinement***

Using the three-dimensional electron density map obtained from above, the polypeptide chain of the cathepsin K can be traced without ambiguity. All 215 residues with side chains were built using the 3-D computer graphics program FRODO [Jones, T.A., *J. Appl. Crystallogr.*, **11**, 268-272 (1978)]. Each of these 215 amino acids residues was manually positioned in its electron density, allowing for a unique position for each atom in cathepsin K in which each position is defined by a unique set of atomic coordinates (X,Y,Z) as shown in Table I. Starting with these atomic coordinates, a diffraction pattern was calculated and compared to the experimental data. The difference between the calculated and experimentally determined diffraction patterns was monitored by the value of R_c. The refinement (using X-PLOR) of the structural model necessitates adjustments of atomic positions to minimize the R-factor, where a value of below 20% is typical for a good quality protein structure and a value of higher than 25% usually indicates the need of further refinement.

<u>EXAMPLE 2: Assays</u>

Determination of cathepsin K proteolytic catalytic activity

All assays for cathepsin K were carried out with human recombinant enzyme. Standard assay conditions for the determination of kinetic constants used a fluorogenic peptide substrate, typically Cbz-Phe-Arg-AMC, and were determined in 100 mM Na acetate at pH 5.5 containing 20 mM cysteine and 5 mM EDTA. Stock substrate solutions were prepared at concentrations of 10 or 20 mM in DMSO with 20 uM final substrate concentration in the assays. All assays contained 10% DMSO. Independent experiments found that this level of DMSO had no effect on enzyme activity or kinetic constants. All assays were conducted at ambient temperature.

Product fluorescence (excitation at 360 nM; emission at 460 nM) was monitored with a Perceptive Biosystems Cytofluor II fluorescent plate reader. Product progress curves were generated over 20 to 30 minutes following formation of AMC product.

Inhibition studies

Potential inhibitors were evaluated using the progress curve method. Assays were carried out in the presence of variable concentrations of test compound. Reactions were initiated by addition of enzyme to buffered solutions of inhibitor and substrate. Data analysis was conducted according to one of two procedures depending on the appearance of the progress curves in the presence of inhibitors. For those compounds whose progress curves were linear, apparent inhibition constants ($K_{i,app}$) were calculated according to equation 1 (Brandt *et al.*, *Biochemistry*, **1989**, *28*, 140):

(1) $$v = V_m A / [K_a(1 + I/K_{i,\,app}) + A]$$

where v is the velocity of the reaction with maximal velocity V_m, A is the concentration of substrate with Michaelis constant of K_a, and I is the concentration of inhibitor.

For those compounds whose progress curves showed downward curvature characteristic of time-dependent inhibition, the data from individual sets was analyzed to give k_{obs} according to equation 2:

(2) $$[\mathrm{AMC}] = v_{ss}\, t + (v_0 - v_{ss})\, [1 - \exp(-k_{obs} t)] / k_{obs}$$

where [AMC] is the concentration of product formed over time t, v_0 is the initial reaction velocity and v_{ss} is the final steady state rate. Values for k_{obs} were then analyzed as a linear function of inhibitor concentration to generate an apparent second order rate constant (k_{obs} / inhibitor concentration or k_{obs} / [I]) describing the time-dependent inhibition. A complete discussion of this kinetic treatment has been fully described (Morrison *et al.*, *Adv. Enzymol. Relat. Areas Mol. Biol.*, **1988**, *61*, 201).

This assay measures the affinity of inhibitors to cathepsin K. One skilled in the art would consider any compound exhibiting a K_i value of less than 50 micromolar to be a potential lead compound for further research. Preferably, the compounds used in the method of the present invention have a K_i value of less than 1 micromolar. Most preferably, said compounds have a K_i value of less than 100 nanomolar.

Human Osteoclast Resorption Assay

Aliquots of osteoclastoma-derived cell suspensions were removed from liquid nitrogen storage, warmed rapidly at 37°C and washed x1 in RPMI-1640 medium by centrifugation (1000 rpm, 5 min at 4°C). The medium was aspirated and replaced with murine anti-HLA-DR antibody, diluted 1:3 in RPMI-1640 medium, and incubated for 30 min on ice The cell suspension was mixed frequently.

The cells were washed x2 with cold RPMI-1640 by centrifugation (1000 rpm, 5 min at 4°C) and then transferred to a sterile 15 mL centrifuge tube. The number of mononuclear cells were enumerated in an improved Neubauer counting chamber.

Sufficient magnetic beads (5 / mononuclear cell), coated with goat anti-mouse IgG, were removed from their stock bottle and placed into 5 mL of fresh medium (this washes away the toxic azide preservative). The medium was removed by immobilizing the beads on a magnet and is replaced with fresh medium.

The beads were mixed with the cells and the suspension was incubated for 30 min on ice. The suspension was mixed frequently. The bead-coated cells were immobilized on a magnet and the remaining cells (osteoclast-rich fraction) were decanted into a sterile 50 mL centrifuge tube. Fresh medium was added to the bead-coated cells to dislodge any trapped osteoclasts. This wash process was repeated x10. The bead-coated cells were discarded.

The osteoclasts were enumerated in a counting chamber, using a large-bore disposable plastic Pasteur pipette to charge the chamber with the sample. The cells were pelleted by centrifugation and the density of osteoclasts adjusted to 1.5×10^4/mL in EMEM medium, supplemented with 10% fetal calf serum and 1.7g/liter of sodium bicarbonate. 3 mL aliquots of the cell suspension (per treatment) were decanted into 15 mL centrifuge tubes. These cells were pelleted by centrifugation. To each tube 3 mL of the appropriate treatment was added (diluted to 50 uM in the EMEM medium). Also included were appropriate vehicle controls, a

positive control (87MEM1 diluted to 100 ug/mL) and an isotype control (IgG2a diluted to 100 ug/mL). The tubes were incubate at 37°C for 30 min.

0.5 mL aliquots of the cells were seeded onto sterile dentine slices in a 48-well plate and incubated at 37°C for 2 h. Each treatment was screened in quadruplicate. The slices were washed in six changes of warm PBS (10 mL / well in a 6-well plate) and then placed into fresh treatment or control and incubated at 37°C for 48 h. The slices were then washed in phosphate buffered saline and fixed in 2% glutaraldehyde (in 0.2M sodium cacodylate) for 5 min., following which they were washed in water and incubated in buffer for 5 min at 37°C. The slices were then washed in cold water and incubated in cold acetate buffer / fast red garnet for 5 min at 4°C. Excess buffer was aspirated, and the slices were air dried following a wash in water.

The TRAP positive osteoclasts were enumerated by bright-field microscopy and were then removed from the surface of the dentine by sonication. Pit volumes were determined using the Nikon/Lasertec ILM21W confocal microscope.

EXAMPLE 3: Method of Detecting Inhibitors

The three dimensional atomic structure can be readily used as a template for selecting potent inhibitors. Various computer programs and databases are available for the purpose. A good inhibitor should at least have excellent steric and electrostatic complementarity to the target, a fair amount of hydrophobic surface buried and sufficient conformational rigidity to minimize entropy loss upon binding. The approach usually comprises several steps:

1) Define a region to target. the active site cavity of cathepsin K can be selected, but any place that is essential to the protease activity could become a potential target. Since the crystal structure has been determined, the spatial and chemical properties of the target region is known.

2) Docking a small molecule onto the target. Many methods can be used to archive this. Computer databases of three-dimensional structures are available for screening millions of small molecular compounds. A negative image of these compounds can be calculated and used to match the shape of the target cavity. The profiles of hydrogen bond donor-acceptor and lipophilic points of these compounds can also be used to complement those of the target. Anyone skilled in the art would be able to identify many small molecules or fragments as hits.

3) Linking and extending recognition fragments. Using the hits identified by above procedure, one can incorporate different functional groups or small molecules into a single, larger molecule. The resulting molecule is likely to be more potent and have higher specificity. It is also possible to try to improve the "seed" inhibitor by adding more atoms or fragments that will interact with the target protein. The originally defined target region can be readily expanded to allow further necessary extension.

A limited number of promising compounds can be selected through the process. They can then be synthesized and assayed for their inhibitory properties. The success rate can sometimes be as high as 20%, and it may still be higher with the rapid progresses in computing methods.

EXAMPLE 4: Crystallization of Enzyme with Inhibitors

A. Preparation of Inhibitors

Compound 1. Preparation of 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

a) 3-hydroxy-4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-pyrrolidinecarboxylic acid 1,1dimethylethyl ester

To a solution of 3-hydroxy-4-amino-1-pyrrolidinecarboxylic acid, 1,1-dimethylethyl ester (202 mg, 1.14 mmol) in CH_2Cl_2 (5 mL) was added CBZ-leucine (302.9 mg, 1.14 mmol), HOBT (154 mg, 1.14 mmol) and EDC (262.2 mg, 1.37 mmol). The reaction was allowed to stir until complete by TLC analysis whereupon it was diluted with EtOAc and washed sequentially with pH 4 buffer, sat. K_2CO_3, water and brine. The organic layer was dried ($MgSO_4$), filtered and concentrated. Column chromatography of the residue (3:1 EtOAc:hexanes) gave 325 mg of the title compound: MS (ES+) 450.3 (MH+), 472.2 (M+Na).

b) 3-hydroxy-4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-pyrrolidine hydrochloride

To a solution of the carbamate (310 mg, 0.69 mmol) in dry EtOAc (5.0 mL) was bubbled HCl gas for approximately 5 minutes. The reaction was stirred until TLC analysis indicated the complete consumption of the starting material. The

reaction was then concentrated *in vacuo* to give 249 mg of the title compound: MS (ES+) 350.3 (MH+)

c) 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinol

To a solution of the amine hydrochloride from the previous step (249 mg, 0.64 mmol) in CH_2Cl_2 (10 mL) was added CBZ-leucine (170.4 mg, 0.64 mmol), HOBT (86.5 mg, 0.64 mmol), NMM (300 uL) and EDC (147.2 mg, 0.77 mmol). The reaction was allowed to stir at room temperature for 2 hours whereupon it was diluted with ethyl acetate and worked up as described previously. Column chromatography of the residue (3:1EtOAc:hexanes) gave 104 mg of the title compound: MS (ES+) 597.1 (MH+), 619.1 (M+Na).

d) 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

To a 0^oC solution of the alcohol (100 mg, 0.17 mmol) in acetone (5.0 mL) was added Jone's reagent dropwise until the brown color persisted. The reaction was allowed to warm to room temperature and stirred approximately 48 hours whereupon it was quenched with isopropanol, diluted with EtOAc and washed sequentially with sat. K_2CO_3, water and brine. The organic layer was dried ($MgSO_4$), filtered and concentrated. Column chromatography of the residue (3:1 EtOAc:hexanes) gave 31 mg of the title compound: MS (ES+) 595.1 (MH+), 617.0 (M+Na).

Compound 2. Preparation of 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone

a) 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(tert-butoxy)carbonyl]-N-(methyl)-L-leucyl]-3-pyrrolidinol

To a solution of 3-hydroxy-4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-pyrrolidine (350 mg) was added N-BOC-N-methyl-leucine (222 mg, 0.0.91 mmol), HOBT(122.5 mg, 0.91 mmol), EDC (208.6 mg, 1.08 mmol) and N-methyl morpholine (0.3 mL, 2.72 mmol). The reaction was stirred at room temperature until complete by TLC analysis. Workup and column chromatography (1:1 Hex:EtOAc) gave 480 mg of the title compound which was used in the following reaction: MS (ES+) 477.4, 577.4 (MH+), 599.4 (M+Na).

b) 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(tert-butoxy)carbonyl]-N-(methyl)-L-leucyl]-3-pyrrolidinone

To a -78°C solution of oxalyl chloride (0.11 mL, 1.23 mmol) in CH_2Cl_2 was added DMSO (0.17 mL, 2.46 mmol) dropwise. The reaction was allowed to stir at -78°C for 20 minutes whereupon a solution of the alcohol (474 mg, 0.82 mmol) in CH_2Cl_2 was added dropwise. The reaction was stirred at -78°C for 30 minutes whereupon triethylamine (0.57 mL) was added in a single portion and allowed to warm to room temperature. Workup and column chromatography (2:1 hexanes:ethyl acetate) gave 247 mg of the title compound: MS (ES+) 475, 575 (M+H), 597 (M+Na).

c) 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone hydrochloride

To a room temperature solution EtOAc/HCl was added the carbamate. The reaction was stirred until complete by TLC analysis. Concentration gave the title compound: MS (ES+) 475 (M+H, 100%).

<u>Compound 3. Preparation of 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone</u>

a) 3-hydroxy-4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-pyrrolidinecarboxylic acid 1,1dimethylethyl ester

3-hydroxy-4-amino-1-pyrrolidinecarboxylic acid, 1,1-dimethylethyl ester was coupled with iso-nicotinoyloxycarbonyl leucine in a similar manner as that described above to give 8.5 grams of the title compound: MS (ES+) 451 (MH+, 100%).

b) 3-hydroxy-4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-pyrrolidine hydrochloride

The carbamate from the previous step was deprotected with EtOAc/HCl to give 8.4 grams of the title compound after concentration: MS (ES+)351 (MH+, 100%).

c) 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinol

To a solution of CBZ leucinal (155 mg) in CH_2Cl_2 was added triethylamine (0.09 mL) and the amine hydrochloride (200 mg, 0.52 mmol) from the previous step. The reaction was stirred at room temperature for 2 hours whereupon the majority of the solvent was removed *in vacuo*. The mixture was redissolved in CH_2Cl_2 and sodium triacetoxyborohydride was added. The reaction was stirred at room temperature for 4 hours. Workup and column chromatography (5% methanol/chloroform) gave 200.5 mg of the title compound: MS(ES+) 583 (MH+, 100%).

d) 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

To a DMSO (2 mL) solution of the alcohol (50 mg, 0.09 mmol) from the previous step was added triethylamine (0.07 mL, 0.52 mmol) and pyridine/sulfur trioxide complex (41 mg, 0.26 mmol). The reaction was maintained at room temperature until complete by TLC analysis. Workup and chromatography (5% methanol/chloroform) gave 37 mg of the title compound: MS (ES+) 582 (MH+, 100%).

<u>Compound 4. Preparation of (3S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-1-(1-propoxy)-5-methyl-2-hexanone</u>

(3S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-1-diazo-5-methyl-2-hexanone (150 mg, 0.37 mmol) was dissolved in 1-propanol (2.5 ml), then rhodium acetate (2 mg) was added and the reaction was stirred at RT for 2h. The reaction mixture was chromatographed (silica gel, 20% EtOAc/hexanes) to yield the title compound as a white solid (59 mg, 37%). MS(ES) $M+H^+ = 435$, $M+NH_4^+ = 452$, $2M+H^+ = 869.6$.

<u>Compound 5. Preparation of bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one</u>

Cbz-leucine (500 mg, 1.88 mmol), EDCI (558 mg, 1.88 mmol) was dissolved in DMF (4.0 ml) with 1,3-diamino-propan-2-ol (85 mg, 0.94 mmol) and Hunig's base (0.3 ml, 1.88 mmol) and was stirred at RT overnight. The reaction was diluted with EtOAc (20 ml) and was extracted with water (2 x 20 ml). The combined organics were dried with magnesium sulfate, filtered, concentrated in vacuo. The intermediate was then dissolved in acetone (4.0 ml) and Jones reagent

(2.0 ml, 1.5 M) was added dropwise and the reaction was stirred at RT overnight. The excess Jones reagent was then quenched with isopropanol (1.0 ml), then the reaction was diluted with EtOAc (20 ml) and was extracted with water (2x 20 ml) to remove the inorganic salts. The combined organics were dried with magnesium sulfate, filtered, concentrated, and chromatographed (silica gel, 2-5% MeOH/ methylene chloride) to give the title compound as a white solid (410 mg, 75%). MS(ES) $M+H^{+}=583$, $M+Na^{+}=605$.

Compound 6. Preparation of 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide

a) methyl 3-benzyloxybenzoate

To a suspension of NaH (0.395 g, 9.87 mmol, 60% in mineral oil) in DMF (20 mL) was added methyl 3-hydroxybenzoate (1.0 g, 6.58 mmol). After stirring for 15 min at room temperature, benzyl bromide (1.1 g, 6.58 mmol) was added. After stirring at room temperature for 3h, the solution was partitioned between ethyl acetate and water. The organic layer was washed with water (2 X 75 mL), saturated aqueous sodium bicarbonate, and brine, then dried ($MgSO_4$), filtered and concentrated to yield an off-white solid (1.013 g, 4.2 mmol). ^1H NMR (400 MHz, $CDCl_3$) d 7.67 (m, 2H), 7.48-7.34 (m. 6H), 7.19 (m, 1H), 5.12 (s, 2H), 3.95 (s, 3H).

b) 3-benzyloxybenzoic acid

To a solution of the compound of Example 6(a) (0.400 g, 1.65 mmol) in THF (2 mL) and water (2 mL) was added lithium hydroxide monohydrate (0.076 g, 1.82 mmol). After stirring at reflux for 5 h, the solution was partitioned between ethyl acetate and 3N HCl. The organic layer was washed with brine, dried ($MgSO_4$), filtered and concentrated to yield a white solid (0.355 g, 1.56 mmol). ^1H NMR (400 MHz, CD_3OD) d 7.58 (m, 2H), 7.36-7.24 (m. 6H), 7.10 (m, 1H), 5.04 (s, 2H).

c) 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide

Following the procedure of Example A, below, except substituting 3-benzyloxybenzoic acid for N-acetyl-L-leucine and 2-[N-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide for 2-[N-(N-benzyloxycarbonyl-L-alanyl)]carbohydrazide,

the title compound was prepared as a white solid (0.062 g, 25%). MS(ESI): 548.1 (M+H)+.

Example A

Preparation of 2-[N-(N-acetyl-L-leucinyl)]-2'-[N'-(N-benzyloxycarbonyl-L-alanyl)]carbohydrazide

To a stirring solution of 2-[N-(N-benzyloxycarbonyl-L-alanyl)]carbohydrazide (0.150g, 0.508mmol) in DMF (2mL) was added N-acetyl-L-leucine (0.092g, 0.534mmol), 1-hydroxybenzotriazole (0.014g, 0.102mmol), and 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (0.102g, 0.534mmol). After stirring at room temperature for 16h, the solution was diluted with ethyl acetate, washed successively with water, saturated aqueous sodium bicarbonate, and brine. The organic layer was dried (MgSO4), filtered and concentrated. The residue was purified by column chromatography (silica gel, methanol/dichloromethane) to yield the title compound as a white solid (0.028 g, 12%). MS(ESI): 451.1 (M+H)+.

Compound 7. Preparation of (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide

a) N-*tert*-butoxycarbonyl-(L)-leucinamide

To a solution of N-tert-butoxycarbonyl-(L)-leucine (7.0g, 28.1mmol) in dry THF (100mL) at -40°C was added isobutylchloroformate (3.8g, 28.1mmol) and N-methylmorphiline (6.0, 59mmol). After 15 minutes of stirring, ammonia was bubbled through the mixture for an additional 15 minutes, then warmed to room temperature and allowed to stir for 2 hours. Mixture filtered and filtrate concentrated in vacuo to yield title compound as a white solid (6.5, 28.0mmol). ^{1}HNMR (400MHz, $CDCl_3$) d 6.38 (br s, 1H), 5.79 (br s, 1H), 5.04 (br d, 1H), 4.13 (m, 1H), 1.71-1.49 (m, 3H), 1.39 (s, 9H), 0.92 (dd, 6H).

b) N-*tert*-butoxycarbonyl-(L)-leucinethioamide

To a stirring solution of the compound of Example 7(a) (6.5, 28.0 mmol) in dry THF was added Lawesson's reagent (6.8g, 16.9 mmol) and the mixture was stirred at room temperature under argon overnight. The solvent was evaporated and the residue chromatographed (silica gel, 12% ethyl acetate/hexane) to give the title compound as a white solid (5.4g, 77%). ^{1}HNMR (400MHz, $CDCl_3$) d 8.54 (br s,

1H), 7.97 (br s, 1H), 5.28 (br d, 1H), 4.52 (m, 1H), 1.72-1.58 (m, 3H), 1.40 (s, 9H), 0.92 (m, 6H).

c) (1S)-1-(tert-butoxycarbonyl)amino-1-(4-carboethoxythiazol-2-yl)-3-methylbutane

The compound of Example 7(b) (5.4g, 21.7 mmol) was stirred in dry acetone (100mL) under argon at -10°C. Ethylbromopyruvate (4.7g, 23.9mmol) was added and stirred for 1h at -10°C. The solution was poured into a well stirred mixture of chloroform and water and then into saturated sodium bicarbonate solution. The organic phase was separated and the aqueous layer extracted with chloroform. The combined organic extracts were dried over $MgSO_4$, filtered and concentrated to an oil. The oily residue was treated with TFAA (5.0g, 23.9mmol) and pyridine (3.8g, 47.8mmol) in dichloromethane for 1h at -20°C. Excess solvent was removed in vacuo and the residue was dissolved in dichloromethane. The solution was washed with saturated aqueous sodium bicarbonate and 1.0N $KHSO_4$ until pH 7. The solution was dried over magnesium sulfate, filtered and concentrated to an oil which was chromatographed (silica gel, 7.5% ethyl acetate/hexane) to give the title compound as a tan solid (4.5g, 61%). ^{1}HNMR (400MHz, $CDCl_3$) d 7.98 (s, 1H), 5.04 (br d, 1H), 4.95 (m, 1H), 4.31 (q, 2H), 1.88 (m, 1H), 1.63 (m, 2H), 1.40 (s, 9H),1.32 (t, 3H), 0.85 (dd, 6H).

d) (1S)-1-(Benzyloxycarbonyl)amino-1-(4-carboethoxythiazol-2-yl)-3-methylbutane

The compound of Example 7(c) (0.250g, 0.731mmol) was dissolved in TFA (2mL) and stirred at room temperature for 15 minutes when diluted with methanol and concentrated in vacuo. The residue was dissolved in methylene chloride and treated with triethylamine (0.739g, 7.31mmol) followed by benzyl chloroformate (1.2g, 7.31mmol). The solution stirred at room temperature for 2h when partition between ethyl acetate/water. The organic layer was washed with brine, collected, dried ($MgSO_4$) and concentrated to a residue that was chromatographed (silica gel, 15% ethyl acetate/hexane) to give the title compound as an oil (0.198g, 72%). ^{1}HNMR (400MHz, $CDCl_3$) d 8.01 (s, 1H), 7.32 (m, 5H), 5.51 (br d, 1H), 5.14 (m, 1H), 5.10 (s, 2H), 4.37 (q, 2H), 1.93 (m, 1H), 1.81-1.67 (m, 2H), 1.39 (t, 3H), 0.95 (m, 6H).

e) (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide

Following the procedure of Example B(a)-(d), below, except substituting (1S)-1-(Benzyloxycarbonyl)amino-1-(4-carboethoxythiazol-2-yl)-3-methylbutane for (1S)-1-benzyloxycarbonylamino-1-(2-carboethoxythiazol-4-yl)-3-methylbutane in step (c), the title compound was prepared. MS (MH^+): 610.0

Example B

Preparation of (1S,2'R)-N-4-[[(1-benzyloxycarbonyl)amino]-3-methylbutyl]thiazol-2-ylcarbonyl-N'-2'-(benzyloxycarbonyl)amino-4'-methylpentanoylhydrazide

a) N-benzyloxycarbonyl-L-leucinyl bromomethyl ketone

1-methyl-3-nitro-1-nitrosoguanidine (6.65 g, 45.2 mmol) in ether (225 mL) is cooled to 0°C. 40% sodium hydroxide is added slowly and the diazomethane is allowed to collect in the ether solution for 30 minutes at 0°C. The ether solution is then decanted and left at 0 °C.

N-Cbz-L-leucine (2.10 g, 7.6 mmol) was dissolved in THF (10 mL), cooled to -40 °C, and 4-methylmorpholine (0.77 g, 7.6 mmol, 0.83 mL) was added, followed by dropwise addition of isobutyl chloroformate (1.04 g, 7.6 mmol, 0.98 mL). After 15 min, the solution was filtered into the previously prepared 0 °C solution of ethereal diazomethane. The resulting solution was allowed to stand at 0 °C for 23 h. HBr (30% in acetic acid) (45.2 mmol, 9 mL) was added and the resulting solution was stirred at 0 °C for 5 min, then washed sequentially with 0.1 N HCl, saturated aqueous NaHCO3 and saturated brine, then dried (MgSO4), filtered and concentrated to give the title compound as a colorless oil (2.43 g, 94%).

b) (1S)-1-benzyloxycarbonylamino-1-(2-carboethoxythiazol-4-yl)-3-methylbutane

A solution of the compound of Example B(a) (1.57 g, 4.58 mmol) and ethyl thiooxamate (0.61 g, 4.58 mmol) in ethanol (10 mL) was heated at reflux for 4 h. The solution was then cooled , concentrated and the residue was purified by flash chromatography on 230-400 mesh silica gel, eluting with 1:4 ethyl acetate/hexanes, to give the title compound as a yellow oil (1.0 g, 58%). 1H NMR (400 MHz, CDCl3) d 7.41 (s, 1H), 7.34-7.31 (m, 5H), 5.40 (d, 1H), 5.10 (d, 1H), 5.05 (d, 1H),

4.98 (q, 1H), 4.48 (q, 2H), 1.80-1.76 (m, 2H), 1.57-1.53 (m, 1H), 1.44 (t, 3H), 0.95 (d, 3H), 0.93 (d, 3H).

c) (1S)-1-benzyloxycarbonylamino-1-(2-hydrazinocarbonylthiazol-4-yl)-3-methylbutane

A solution of the compound of Example B(b) (0.30 g, 0.8 mmol) and hydrazine hydrate (0.40 g, 8.0 mmol, 0.39 mL) in ethanol (8 mL) was allowed to stir at room temperature for 2 h. The solution was then concentrated to yield the title compound as a white foam (0.28 g, 98%). 1H NMR (400 MHz, CDCl3) d 8.29 (s, 1H), 7.37-7.35 (m, 5H), 5.18 (d, 1H), 5.09 (dd, 2H), 4.95 (q, 1H), 4.07 (d, 2H), 1.71 (t, 2H), 1.55 (m, 1H), 0.96 (d, 3H), 0.94 (d, 3H).

d) (1S,2'R)-N-4-[[(1-benzyloxycarbonyl)amino]-3-methylbutyl]thiazol-2-ylcarbonyl-N'-2'-(benzyloxycarbonyl)amino-4'-methylpentanoylhydrazide

A solution of the compound of Example B(c) (100 mg, 0.28 mmol), N-Cbz-L-leucine (80.5 mg, 0.30 mmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (58.2 mg, 0.30 mmol) and 1-hydroxybenzotriazole (7.5 mg, 0.06 mmol) in DMF (0.6 mmol) was allowed to stir at room temperature for 18 h. The solution was diluted with ethyl acetate and washed successively with water, 0.1 N HCl, saturated aqueous NaHCO3 and saturated brine, then dried (MgSO4), filtered and concentrated. The residue was purified by flash chromatography on 230-400 mesh silica gel, eluting with 1:1 ethyl acetate/hexanes, to provide the title compound as a white solid (111.4 mg, 66%). mp 110-112 °C.

Compound 8. Preparation of 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide

To a stirring solution of N-Cbz-L-leucine (Chemical Dynamics Corp.) (2.94 g, 11.1 mmol) in 22 mL of DMF was added carbohydrazide (0.5 g, 5.6 mmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (2.13 g, 11.1 mmol) and 1-hydroxybenzotriazole (0.3 g, 2.2 mmol). After stirring at room temperature for 22 h, the solution was poured into 500 mL of water. The precipitate was collected by vacuum filtration and washed with water (4 X 150 mL) and dichloromethane (4 X 150 mL), then dried under vacuum to provide the title compound as a white solid (1.49 g, 46%). MS(ESI): 607.1 $(M+Na)^+$.

Compound 9. Preparation of 1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one

a) 1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy phenyl sulfonyl)-amino-propan-2-one

Following the procedure of Example C(a)-(d), below, substituting "imidazole acetic acid" for "4-pyridyl acetic acid", the title compound was prepared: MS(ES) $M+H^+ = 542$.

Example C

Preparation of 1-N-(N-Cbz-leucinyl)-amino-3-N-(2-pyridyl-sulfonyl)-amino-propan-2-one

a) 1-N-(N-Cbz-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-ol

1,3-Diamino propan-2-ol (6.75 g, 75 mmol) was dissolved in DMF (100ml) and Cbz-leucine (20g, 75.5 mmol), HOBT-hydrate (11g, 81.5 mmol), and EDCI (15.5g, 81.2 mmol) were added. The reaction was stirred overnight at RT. A portion of the reaction mixture (30 ml) was concentrated in vacuo, then ether (50 ml) and MeOH (30 ml) were added. A 1N solution of hydrochloric acid in ether was added (1 M, 30 ml) and a white gum formed, which was washed several times with ether. MeOH-acetone were added and heated until the gum became a white solid. The white solid was dissolved in DMF (25 ml) and DIEA (5ml), then 4-phenoxy

phenyl sulfonyl chloride was added. The reaction was stirred for 2h, concentrated in vacuo, then chromatographed (silica gel, 1:1 EtOAc: hexanes) to provide the desired product as a white solid.

b) Leucinyl-amino-3-N-(4-phenoxy phenyl sulfonyl)-amino-propan-2-ol

1-N-(Cbz-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-ol (1.0g, 1.8 mmol) was dissolved in EtOH (30 ml), then 10% Pd/C (0.22g) was added followed by 6N hydrochloric acid (2.5 ml), and the reaction was stirred under a balloon of hydrogen gas for 4h at RT. The reaction mixture was filtered, concentrated, and azeotroped with toluene to provide a white glass which was used in the next reaction without further purification.

c) 1-N-(N-4-pyridyl acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-ol

Leucinyl-amino-3-N-(4-phenoxy phenyl sulfonyl)-amino-propan-2-ol (0.36 g, 0.76 mmol) was dissolved in DMF (5 ml), then NMM (0.45 ml, 4 mmol) was added followed by 4-pyridyl acetic acid (0.13g, 0.75 mmol) and HBTU (0.29g, 0.76 mmol) and the reaction was stirred at RT overnight. The reaction mixture was concentrated in vacuo, then chromatographed (silica gel, 5%MeOH: methylene chloride) to provide the desired product as a white solid (90 mg, MS(ES): $M+H^+$ = 555.

d) 1-N-(N-4-pyridyl acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one

1-N-(N-4-pyridyl-acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-ol (45 mg, 0.08 mmol) was dissolved in acetone (5ml), then 1N hydrochloric acid (2 ml) was added. The reaction was concentrated in vacuo, then redissolved in acetone. Jones reagent (1.5 M, several drops) was added and the reaction mixture was stirred for 6h at RT. Isopropanol (0.5 ml) was added and the reaction mixture was concentrated in vacuo. The reaction was diluted with pH 7 buffer and then was extracted with EtOAc, dried with magnesium sulfate, filtered, concentrated in vacuo, then chromatographed (silica gel, 5% MeOH-methylene chloride) to give the desired product as a white solid (27 mg, 50%): MS(ES): $M+H^+$ = 553.

B. Crystallization of the protein and protein-inhibitor complexes

Human cathepsin K was expressed in *baculovirus* cells for the first eight of the nine inhibitors described below. Conditioned media containing expressed pro-cathepsin K was loaded directly onto an S-Sepharose column pre-equilibrated with 25 mM phosphate buffer at pH 8. The column was eluted with a NaCl gradient. Fractions containing pro-cathepsin K were pooled, concentrated to 2.5 mg/ml and activated to mature cathepsin K in 50 mM sodium acetate buffer pH 4.0 containing 20 mM L-cysteine and 1% mature cathepsin K as seed. The activation was monitored using CBZ-Phe-Arg-AMC,as fluorogenic substrate and by SDS-PAGE. When the increasing specific activity reached a plateau (*ca.* 15 μmol/min/mg), the reaction was stopped by the addition of inhibitor. The inhibited mature cathepsin K was concentrated and dialyzed against 20 mM MES, 50 mM NaCl, 2 mM L-cysteine, pH 6.

Protein preparation for cathepsin K complex with 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone (only)

Human cathepsin K was expressed in *E. coli*. The cell pellet from 1 L of bacterial culture weighing 2.35 gm. was washed with 50 mL of 50 mM Tris/HCl, 5 mM EDTA, 150 mM NaCl, pH 8.0. After centrifugation at 13,000 x g for 15 mins, the washed pellet was resuspended into 25 mL of the same buffer prepared at 4° C and lysed by passage twice through a cell disruptor (Avestin) at 10,000 psi. The lysate was centrifuged as above, the supernatant decanted and the pellet suspended in 25 mL 50 mM Tris/HCl, 10 mM DTT, 5 mM EDTA, 150 mM NaCl, pH 8.0 containing either 8 M urea or 6 M guanidine HCl. After stirring at 4° C for 30 mins, insoluble cellular debris was removed by centrifugation at 23,000 x g for 30 mins and the supernatant clarified by filtration (0.45 um, Millipore).

Varying amounts of the proenzyme form of cathepsin K were refolded by quick dilution into stirring, N_2 (g) sparged 50 mM Tris/HCl, 5 mM EDTA, 10 mM reduced and 1 mM oxidized glutathione, 0.7 M L-arginine pH 8.0 and stirred overnight at 4° C. After concentration to ca.1 mg/mL using a stirred cell fitted with a YM-10 membrane (Amicon), the sample was clarified by centrifugation and filtration then dialyzed against 25 mM Na_2PO_4, 1.0 M NaCl, pH 7.0. The dialysate was applied at a LFR= 23 cm/hr to

a 2.6 x 90 cm column of Superdex 75 (Pharmacia) pre-equilibrated in 25 mM Na_2PO_4, 1.0 M NaCl, pH 7.0. The cathepsin K proenzyme was pooled based upon purity as observed on a reduced, SDS-PAGE gel.

Crystals of mature activated cathepsin K complexed with inhibitor grew to a size of approximately 0.2 mm^3 in about six days at 20°C. The concentration of inhibited cathepsin K used in the crystallization was approximately 8 mg./ml. The method of vapor diffusion in hanging drops was used to grow crystals from the solution of cathepsin K - inhibitor complex. The initial crystal structure to be determined was that of cathepsin K in complex with the cysteine protease inhibitor E64. Crystals of mature activated cathepsin K complexed with E-64 grew to a size of approximately 0.2 mm^3 in six days at 20°C. The concentration of E-64-inhibited cathepsin K used in the crystallization was 8 mg/ml. Vapor diffusion was used in hanging drops from a solution of 10% PEG 8000, 0.1 M Na^+/K^+ phosphate at pH 6.2 containing 0.2M NaCl. Crystals of the complex are orthorhombic, space group $P2_12_12_1$, with cell constants of a=38.4, b=50.7, and c=104.9 Ångstroms. This crystal form will be referred to as Form II. The crystals contain one molecule in the asymmetric unit and contain approximately 40% solvent with a Vm value of 2.1 A^3/Dalton. X-ray diffraction data were measured from a single crystal using a Siemens two-dimensional position-sensitive detector on a Siemens rotating anode generate operating a 5 KW. The structure was determined by molecular replacement using X-PLOR. The starting model consisted of all atoms of the main chain of papain and those side chain atoms predicted to be homologous between the two proteins as determined from sequence alignment. The cross rotation function was calculated using x-ray diffraction data from 10 to 4 Å and a radius of integration of 32 Å. The highest peak was 6.0 σ. A translation search was carried out using data from 8 to 3.5 Ångstroms resulting in the highest peak of 12.5 σ.The resulting model gave an R_c factor of 0.488. This model was refined by rigid-body refinement, and the resulting phases were used to calculate Fourier maps with coefficients $|F_o-F_c|$ and $|2F_o-F_c|$, into which the atomic model of cathepsin K was built using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building. The structure was refined using X-PLOR. The electron density for E-64 was clear in the maps. The inhibitor was built into density and several additional cycles of map fitting and refinement were carried out to a final R_c of 0.191.

Crystallization of the complex of cathepsin K with 3(S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 10% isopropanol, 0.1 M $NaPO_4$ / citrate at pH 4.2. Crystals of the complex are tetragonal, space group $P4_32_12$, with cell constants of a=57.6 Å, and c=131.2 Å. This crystal form will be referred to as Form III. Diffraction data were collected as described above. The crystals contain one molecule in the asymmetric unit and contain 36% solvent with a V_m value of 2.3 Å^3/Dalton. The structure was determined by molecular replacement using X-PLOR at 2.5 Ångstroms resolution. The starting model consisted of all protein atoms of the orthorhombic form of cathepsin K-E64 structure. Molecular replacement was carried out as described above for the cathepsin K-E64 structure determination. The model was refined by rigid-body refinement using X-PLOR, and the resulting phases were used to calculate Fourier maps with coefficients $|F_o-F_c|$ and $|2F_o-F_c|$, into which the atomic model of the inhibitor was built using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building. The structure was refined using X-PLOR. Several cycles of map fitting and refinement were carried out to a final R_c of 0.245.

Crystallization of the complex of cathepsin K with 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 22.5% PEG 8000, 0.075 M sodium acetate at pH 4.5 containing 0.15 M Li_2SO_4. Crystals of the complex grew as Form III. Diffraction data were collected as described above. The structure was determined by rigid body refinement with X-PLOR utilizing the previous Form III protein model at 2.4 Ångstroms resolution. Fourier maps with coefficients $|F_o-F_c|$ and $|2F_o-F_c|$ were used to fit the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement (X-PLOR) was used to refine the structure during model building. Several cycles of map fitting and refinement were carried out to a final R_c of 0.237.

Crystallization of the complex of cathepsin K with bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 10% isopropanol, 0.1 M $NaPO_4$ / citrate at pH 4.2. Crystals of the complex grow as Form III. Diffraction data were collected as described above. The structure was determined by rigid body refinement of the previous Form III protein model at 2.6 Ångstroms resolution. Fourier maps with coefficients $|F_o-F_c|$ and $|2F_o-F_c|$ were used to fit the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building. Several cycles of map fitting and refinement were carried out using X-PLOR to a final R_c of 0.210.

Crystallization of the complex of cathepsin K with 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution 18% PEG 8000, 0.6 M sodium acetate at pH 4.5 containing 0.12 M Li_2SO_4. Crystals of the complex grow in Form III. Diffraction data were collected as described above. The structure was determined by rigid body refinement of the previous Form III protein model with X-PLOR at 2.4 Ångstroms resolution. Fourier maps with coefficients $|F_o-F_c|$ and $|2F_o-F_c|$, were used to the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building using X-PLOR. Several cycles of map fitting and refinement were carried out to a final R_c of 0.218.

Crystallization of the complex of cathepsin K with (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 30% MPD, 0.1 M MES at pH 7.0 and 0.1 M tris buffer at pH 7.0. Crystals of the complex are Form II. Diffraction data were collected as described above. The structure was determined by rigid body refinement of the previous Form II protein model with X-PLOR at 2.3 Ångstroms resolution. Fourier maps with

coefficients $|F_O-F_C|$ and $|2F_O-F_C|$, were used to the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building using X-PLOR. Several cycles of map fitting and refinement were carried out to a final R_c of 0.211.

Crystallization of the complex of cathepsin K with 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 33% MPD, 0.1 M MES at pH 7. Crystals of the complex grow as Form II. Diffraction data were collected as described above. The structure was determined by rigid body refinement of the previous Form II protein model with X-PLOR at 2.2 Ångstroms resolution.. Fourier maps with coefficients $|F_O-F_C|$ and $|2F_O-F_C|$, were used to the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building using X-PLOR. Several cycles of map fitting and refinement were carried out to a final R_c of 0.208.

Crystallization of the complex of cathepsin K with 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 28% MPD, 0.1 M MES at pH 7.0 and 0.1 M tris buffer at pH 7.0. Crystals of the complex Form II. Diffraction data were collected as described above. The structure was determined by rigid body refinement of the previous Form II protein model with X-PLOR at 2.3 Ångstroms resolution. Fourier maps with coefficients $|F_O-F_C|$ and $|2F_O-F_C|$, were used to the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building using X-PLOR. Several cycles of map fitting and refinement were carried out to a final R_c of 0.193.

Crystallization of the complex of cathepsin K with 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 30% MPD, 0.1 M MES at pH 7.0 and 0.1 M tris buffer at pH 7.0. Crystals of the complex Form II. Diffraction data were collected as described above. The structure was determined by rigid body refinement of the previous Form II protein model with X-PLOR at 2.2 Ångstroms resolution.. Fourier maps with coefficients $|F_o-F_c|$ and $|2F_o-F_c|$, were used to the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building using X-PLOR. Several cycles of map fitting and refinement were carried out to a final R_c of 0.267.

<u>Crystallization of the complex of cathepsin K with 1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one</u>

Crystals of mature activated cathepsin K complexed with the inhibitor grew from a solution of 18% PEG 8000, 0.6 M sodium acetate at pH 4.5 containing 0.12 M Li_2SO_4. Crystals of the complex are Form III. Diffraction data were collected as described above. The structure was determined by rigid body refinement of the previous Form II protein model at 2.5 Ångstroms resolution.. Fourier maps with coefficients $|F_o-F_c|$ and $|2F_o-F_c|$ were used to fit the atomic model of the inhibitor using the molecular graphics program FRODO. Conventional positional refinement was used to refine the structure during model building. Several cycles of map fitting and refinement were carried out using X-PLOR to a final R_c of 0.246.

<u>Abbreviations</u>

E-64, [1-[N-[(L-3-*trans*-carboxyoxirane-2carbonyl)-L-leucyl]amino]-4-guanidinobutane]
CBZ, benzyloxycarbonyl
AMC, aminomethylcoumarin
MPD, 2 methyl-2,4-pentanediol
PIPES, piperazone-N,N-bis(2-ethanesulfonic acid)
MES, 2-(N-morpholino)-ethanesulfonic acid
tris, tris(hydroxymethyl)-aminomethane
PEG, polyethyleneglycol
M, Molar

$R_C = \Sigma|(F_O - F_C)| / F_O$

F_O = observed structure amplitude

F_C = calculated structure amplitude

EDTA, ethylenediaminetetraacetic acid

DTT, 1,4-dithiothreitol

SDS-PAGE, sodium dodecylsulfate polyacrylamide gel electrophoresis

This invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

The disclosures of the patents, patent applications and publications cited herein are incorporated by reference in their entireties.

WHAT IS CLAIMED IS:

1. A method of inhibiting cathepsin K which comprises administering to a mammal in need thereof a compound that fits spatially into the active site of cathepsin K, said compound comprising any two of the following:

(i) an electrophilic carbon atom that binds to the side chain sulfur atom of cysteine 25 wherein said electrophilic carbon atom is 1.7-4.0Å from said sulfur atom;

(ii) a hydrophobic group that interacts with tryptophan 184 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tryptophan 184 is 4.10-7.10Å;

(iii) a hydrophobic group that interacts with tyrosine 67, methionine 68, alanine 134, leucine 160, and leucine 209, creating a hydrophobic pocket, and has distance ranges between the centroid of said hydrophobic group and the centroids of the side chain atoms of the amino acid residues of said hydrophobic pocket which are tyrosine 67: 4.91- 5.91Å, methionine 68: 5.74-6.74Å, alanine 134: 4.15-5.15Å, leucine 160: 6.18-7.18Å, and leucine 209: 5.71-6.71Å;

(iv) a hydrophobic group that interacts with tyrosine 67 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tyrosine 67 is 4.10-7.10Å;

(v) an amino group with a pKa of less than 7 or an oxygen atom, each of which interacts with a hydrogen atom donated by the amide nitrogen of glycine 66 wherein the distance between these two atoms is 2.7-3.5Å;

(vi) a hydrophobic group that interacts with the main chain atoms of glutamine 21, cysteine 22 and glycine 23 wherein the distance between the centroid of said hydrophobic group and the centroids of glutamine 21 , cysteine 22 and glycine 23 are 3.7-5.4, 4.9-5.7 and 5.4-6.7Å, respectively; or

(vii) a hydrophobic group that interacts with the side chain atoms of glutamine 143 and asparagine 161 and the main chain of alanine 137 and serine 138 wherein the distance between the centroid of the hydrophobic group and the centroids of glutamine 143, asparagine 161, alanine 137, and serine 138 are 7.9-9.6Å, 4.7-5.4Å, 4.2-5.5Å, and 4.6-6.4Å, respectively.

2. A method of inhibiting cathepsin K which comprises administering to a mammal in need thereof a compound that fits spatially into the active site of cathepsin K, said compound comprising any three or more of the following:

(i) an electrophilic carbon atom that binds to the side chain sulfur atom of cysteine 25 wherein said electrophilic carbon atom is 1.7-4.0Å from said sulfur atom;

(ii) a hydrophobic group that interacts with tryptophan 184 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tryptophan 184 is 4.10-7.10Å;

(iii) a hydrophobic group that interacts with tyrosine 67, methionine 68, alanine 134, leucine 160, and leucine 209, creating a hydrophobic pocket, and has distance ranges between the centroid of said hydrophobic group and the centroids of the side chain atoms of the amino acid residues of said hydrophobic pocket which are tyrosine 67: 4.91- 5.91Å, methionine 68: 5.74-6.74Å, alanine 134: 4.15-5.15Å, leucine 160: 6.18-7.18Å, and leucine 209: 5.71-6.71Å;

(iv) a hydrophobic group that interacts with tyrosine 67 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tyrosine 67 is 4.10-7.10Å;

(v) an amino group with a pKa of less than 7 or an oxygen atom, each of which interacts with a hydrogen atom donated by the amide nitrogen of glycine 66 wherein the distance between these two atoms is 2.7-3.5Å;

(vi) a hydrophobic group that interacts with the main chain atoms of glutamine 21, cysteine 22 and glycine 23 wherein the distance between the centroid of said hydrophobic group and the centroids of glutamine 21 , cysteine 22 and glycine 23 are 3.7-5.4, 4.9-5.7 and 5.4-6.7Å, respectively; or

(vii) a hydrophobic group that interacts with the side chain atoms of glutamine 143 and asparagine 161 and the main chain of alanine 137 and serine 138 wherein the distance between the centroid of the hydrophobic group and the centroids of glutamine 143, asparagine 161, alanine 137, and serine 138 are 7.9-9.6Å, 4.7-5.4Å, 4.2-5.5Å, and 4.6-6.4Å, respectively.

3. A method of inhibiting cathepsin K which comprises administering to a mammal in need thereof a compound that fits spatially into the active site of cathepsin K, said compound comprising:

(i) an electrophilic carbon atom that binds to the side chain sulfur atom of cysteine 25 wherein said electrophilic carbon atom is 1.7-4.0Å from said sulfur atom; and

(ii) a hydrophobic group that interacts with tryptophan 184 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tryptophan 184 is 4.10-7.10Å.

4. The method of claim 3 wherein said hydrophobic group that interacts with tryptophan 184 is an aromatic group.

5. The method of claim 4 wherein the centroid of said aromatic group that interacts with tryptophan 184 is 9.24-11.24Å from the centroid of said electrophilic carbon that binds to the side chain sulfur atom of cysteine 25.

6. The method of claim 3 wherein said electrophilic carbon that binds to the side chain sulfur atom of cysteine 25 is a carbonyl carbon.

7. The method of claim 3 wherein the compound further comprises a hydrophobic group that:

has a centroid which is 5.44-6.94Å from said electrophilic carbon;

interacts with tyrosine 67, methionine 68, alanine 134, leucine 160, and leucine 209, creating a hydrophobic pocket; and

has distance ranges between the centroid of said hydrophobic group and the centroids of the side chain atoms of the amino acid residues of said hydrophobic pocket which are tyrosine 67: 4.91- 5.91Å, methionine 68: 5.74-6.74Å, alanine 134: 4.15-5.15Å, leucine 160: 6.18-7.18Å, and leucine 209: 5.71-6.71Å.

8. The method of claim 7 wherein said hydrophobic group that interacts with said hydrophobic pocket is an isobutyl group.

9. The method of claim 3 wherein the compound further comprises a hydrophobic group that interacts with tyrosine 67 wherein the distance between the centroid of said hydrophobic group and the centroid of the side chain atoms of tyrosine 67 is 4.10-7.10Å.

10. The method of claim 9 wherein said hydrophobic group that interacts with tyrosine 67 is an aromatic group.

11. The method of claim 3 wherein the compound further comprises an amino group with a pKa of less than 7 or an oxygen atom, each of which interacts with a hydrogen atom donated by the amide nitrogen of glycine 66 wherein the distance between these two atoms is 2.7-3.5Å.

12. The method of claim 3 wherein the compound further comprises a hydrophobic group that interacts with the main chain atoms of glutamine 21, cysteine 22 and glycine 23 wherein the distance between the centroid of said hydrophobic group and the centroids of glutamine 21 , cysteine 22 and glycine 23 are 3.7-5.4, 4.9-5.7 and 5.4-6.7Å, respectively.

13. The method of claim 12 wherein said hydrophobic group that interacts with glutamine 21, cysteine 22 and glycine 23 is an isobutyl group.

14. The method of claim 3 wherein the compound further comprises a hydrophobic group that interacts with the side chain atoms of glutamine 143 and asparagine 161 and the main chain of alanine 137 and serine 138 wherein the distance between the centroid of the hydrophobic group and the centroids of glutamine 143, asparagine 161, alanine 137, and serine 138 are 7.9-9.6Å, 4.7-5.4Å, 4.2-5.5Å, and 4.6-6.4Å, respectively.

15. The method of claim 1 wherein the compound is:

3(S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone;

4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone;

4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N-[N-(methyl)-L-leucyl)]-3-pyrrolidinone;

4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone;

bis-(Cbz-leucinyl)-1,3-diamino-propan-2-one;

2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide;

(1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide;

1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one; or

2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide;

or a pharmaceutically acceptable salt thereof.

16. A composition comprising cathepsin K in crystalline form.

17. The composition according to claim 16 wherein cathepsin K has an active site cavity formed by the amino acids in Table XXIX.

18. The composition of claim 17 wherein said active site is characterized by the coordinates selected from the group consisting of the coordinates of Tables I-X.

19. A cathepsin K crystal.

20. An isolated, properly folded cathepsin K molecule or fragment thereof having a conformation comprising a catalytically active site formed by the residues listed in Table XXIX, said active site defined by the protein coordinates of Table I.

21. A peptide, peptidomimetic or synthetic molecule which binds with the active site cavity of cathepsin K according to claim 17.

22. A method of identifying an inhibitor compound capable of binding to, and inhibiting the proteolytic activity of, cathepsin K, said method comprising:

introducing into a suitable computer program information defining an active site conformation of a cathepsin K molecule comprising a catalytically active site formed by the residues listed in Table XXIX, said active site defined by the protein coordinates of Table I, wherein said program displays the three-dimensional structure thereof;

creating a three dimensional representation of the active site cavity in said computer program;

displaying and superimposing the model of said test compound on the model of said active site;

assessing whether said test compound model fits spatially into the active site;

preparing said test compound that fits spatially into the active site;

using said test compound in a biological assay for a protease characterized by said active site; and

determining whether said test compound inhibits cathepsin K activity in said assay.

23. A peptide, peptidomimetic or synthetic molecule identified by the method of Claim 22.

24. A method of drug design comprising using the structural coordinates of a cathepsin K crystal to computationally evaluate a chemical entity for associating with the active site of cathepsin K.

25. The method according to claim 24, wherein said entity is a competitive or non-competitive inhibitor of cathepsin K.

26. A method for identifying inhibitors which competitively bind to the active site of a cathepsin K molecule or fragment thereof characterized by a catalytically active site formed by the residues listed in Table XXIX, said method comprising the steps of:

providing the coordinates of said active site of the protease to a computerized modeling system;

identifying compounds which will bind to the structure; and

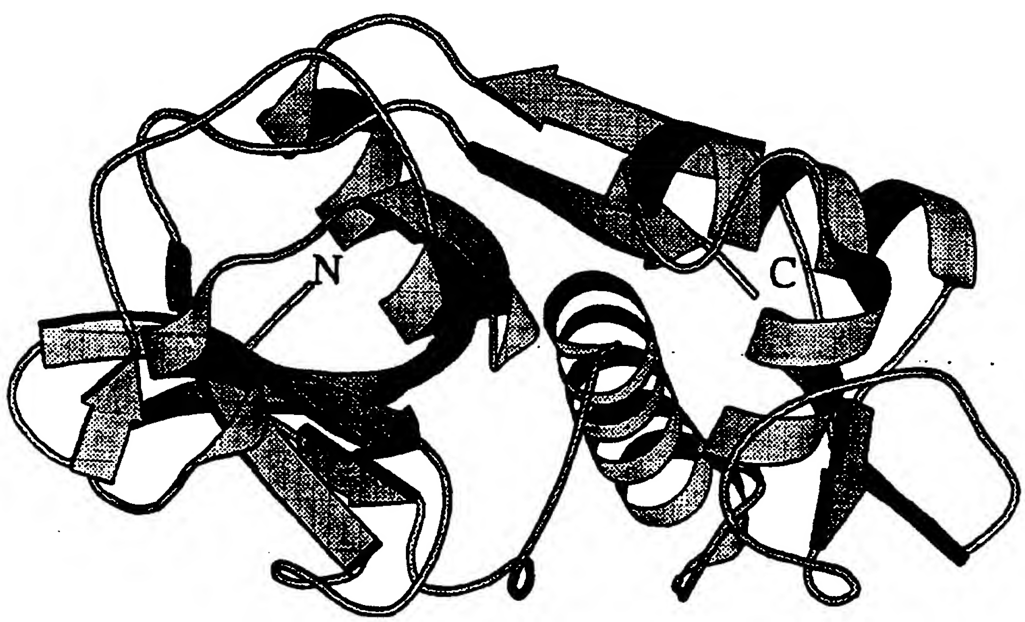
screening the compounds identified for protease inhibitory bioactivity.

FIGURE 1

Sequence Comparison Between Cathepsin K and the Papain Superfamily of Cysteine Proteases

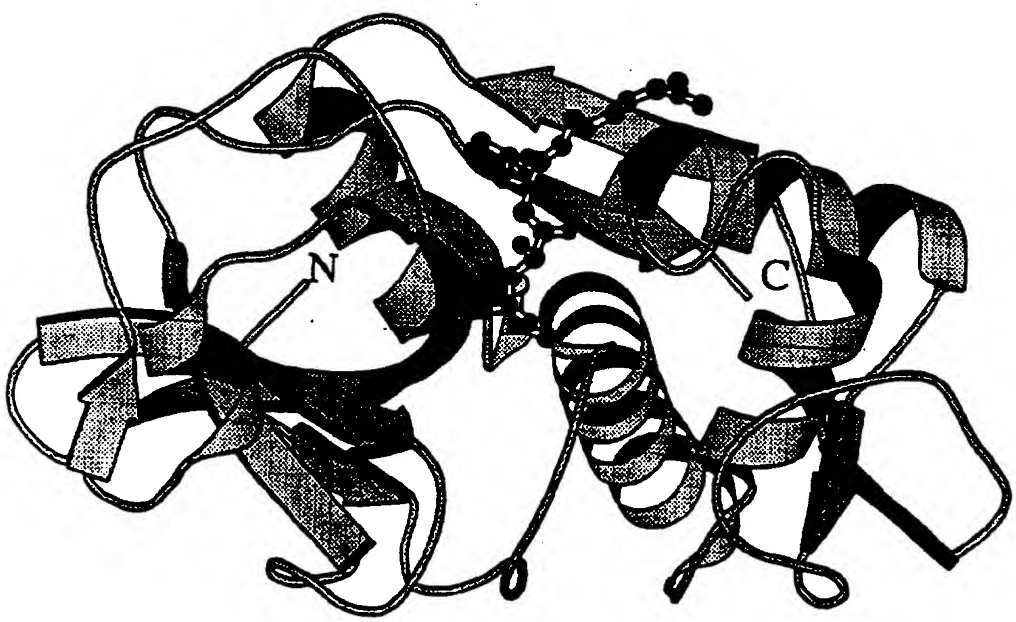
```
           1                                                          50
CatK       .......... .MWGLKVLLL PVVSF....A .LYPEEILDT HWELWKKTHR KQ.YNNKVDE ISRRLIWEKN LKYISIHNLE ASLGVHTYEL
CatS       .......... MKRLVCVLLV CSSAV....A QLHKDPTLDH HWHLWKKTYG KQ.YKEKNEE AVRRLIWEKN LKFVMLHNLE HSMGMHSYDL
CatL       .......... ...MNPTLIL AAFCLGIASA TLTFDHSLEA QWTKWKAMHN RL.Y.GMNEE GWRRAVWEKN MKMIELHNQE YREGKHSFTM
Papain     .......... .......... .......... .......... .......... .......... .......... .......... ..........
Actinidin  MGLPKSFVSM SLLFFSTLLI LSLAFNAKNL TQRTNDEVKA MYESWLIKYG KS.YNSLGEW ERRFEIFKET LRFIDEHNAD ...TNRSYKV
CatH       .......MWA TLPLLCAGAW LLGVPVCGAA ELSVNSLEKF HFKSWMSKHR KT.Y.STEEY HHRLQTFASN WRKINAHN.. ..NGNHTFKM
CatB       .......... .......... .......... .......... MWQLWASLCC LLVLANARSR PSFHPVSDEL VNYVNKRNTT WQAGHNFYNV

           100                                                    150
CatK       AMNHLGDMTS EEVVQKMTGL KVPLSHSRSN DTLYIPEWEG RAPDSVDYRK KG.YVTPVKN QGQCGSCWAF SSVGALEGQL KKKTGKLLN.
CatS       GMNHLGDMTS EEVMSLTSSL RVP.SQWQRN IT.YKSNPNR ILPDSVDWRE KG.CVTEVKY QGSCGACWAF SAVGALEAQL KLKTGKLVT.
CatL       AMNAFGDMTS EEFRQVMNGF Q...NRKPRK GKVFQEPLFY EAPRSVDWRE KG.YVTPVKN QGQCGSCWAF SATGALEGQM FRKTGRLIS.
Papain     .......... .......... .......... .......... .IPEYVDWRQ KG.AVTPVKN QGSCGSCWAF SAVVTIEGII KIRTGNLNQ.
Actinidin  GLNQFADLTD EEFRSTYLGF .TSGSNKTKV SNRYEPRFGQ VLPSYVDWRS AG.AVVDIKS QGECGGCWAF SAIATVEGIN KIVTGVLIS.
CatH       ALNQFSDMSF AEIKHK...Y LWSEPQNCSA TKSNYLRGTG PYPPSVDWRK KGNFVSPVKN QGACGSCWTF STTGALESAI AIATGKMLS.
CatB       DMSYLKRLCG TFLGGPKPPQ RVMFTEDLKL PASFDAR... ......EQWP QCPTIKEIRD QGSCGSCWAF GAVEAISDRI CIHTNAHVSV

                  200                                                    250
CatK       .LSPQNLVDC VSE...NDGC GGGYMTNAFQ YVQKNRGIDS EDAY...... .......... ......PYV GQEESCM... .....YNPTG
CatS       .LSAQNLVDC STEKYGNKGC NGGFMTTAFQ YIIDNKGIDS DASY...... .......... ......PYK AMDQKCQ... .....YDSKY
CatL       .LSEQNLVDC SGPQ.GNEGC NGGLMDYAFQ YVQDNGGLDS EESY...... .......... ......PYE ATEESCK... .....YNPKY
Papain     .YSEQELLDC DR..RSY.GC NGGYPWSALQ LVAQY.GIHY RNTY...... .......... ......PYE GVQRYCR... .....SREKG
Actinidin  .LSEQELIDC GRTQNTR.GC NGGYITDGFQ FIINNGGINT EENY...... .......... ......PYT AQDGECN... .....LDLQN
CatH       .LAEQQLVDC A.QDFNNYGC QGGLPSQAFE YILYNKGIMG EDTY...... .......... ......PYQ GKDGYCK... .....FQP.G
CatB       EVSAEDLLTC CGSMCG.DGC NGGYPAEAWN F.WTRKGLVS GGLYESHVGC RPYSIPPCEH HVNGSRPPCT GEGDTPKCSK ICEPGYSPTY

                      300                                                    350
CatK       K.AAKCRGYR EIPEGNEKAL KRAVARVGPV SVAIDASLTS FQFYSKGVYY DESC..NSDN LNHAVLAVGY G....IQKGN KHWIIKNSWG
CatS       R.AATCSKYT ELPYGREDVL KEAVANKGPV SVGVDARHPS FFLYRSGVYY EPSC...TQN VNHGVLVVGY G....DLNGK EYWLVKNSWG
CatL       S.VANDTGFV DIP.KQEKAL MKAVATVGPI SVAIDAGHES FLFYKEGIYF EPDC..SSED MDHGVLVVGY GFESTESDNN KYWLVKNSWG
Papain     PYAAKTDGVR QVQPYNQGAL LYSIAN.QPV SVVLQAAGKD FQLYRGGIFV GPC....GNK VDHAVAAVGY G....P.... NYILIKNSWG
Actinidin  EKYVTIDTYE NVPYNNEWAL QTAVTY.QPV SVALDAAGDA FKHYSSGIFT GPC....GTA IDHAVTIVGY G....TEGGI DYWIVKNSWD
CatH       KAIGFVKDVA NITIYDEEAM VEAVALYNPV SFAFEVTQD. FMMYRTGIYS STSCHKTPDK VNHAVLAVGY G....EKNGI PYWIVKNSWG
CatB       KQDKHYGYNS YSVSNSEKDI MAEIYKNGPV EGAFSV.YSD FLLYKSGVYQ HVTGEMMG.. .GHAIRILGW G....VENGT PYWLVANSWN

                                 400                           430
CatK       ENWGNKGYIL MARNKNNA.. .CGIANLASF PKM....... .......... .......... ..........
CatS       HNFGEEGYIR MARNKGNH.. .CGIASFPSY PEI....... .......... .......... ..........
CatL       EEWGMGGYVK MAKDRRNH.. .CGIASAASY PTV....... .......... .......... ..........
Papain     TGWGENGYIR IKRGTGNSYG VCGLYTSSFY PVKN...... .......... .......... ..........
Actinidin  TTWGEEGYMR ILRNVGGA.G TCGIATMPSY PVKYNNQNHP KPYSSLINPP AFSMSKDGPV GVDDGQRYSA
CatH       PQWGMNGYFL IERGKN.... MCGLAACASY PIPLV..... .......... .......... ..........
CatB       TDWGDNGFFK ILRGQDHCGI ESEVVAGIPR TDQYWEKI.. .......... .......... ..........
```

FIGURE 2



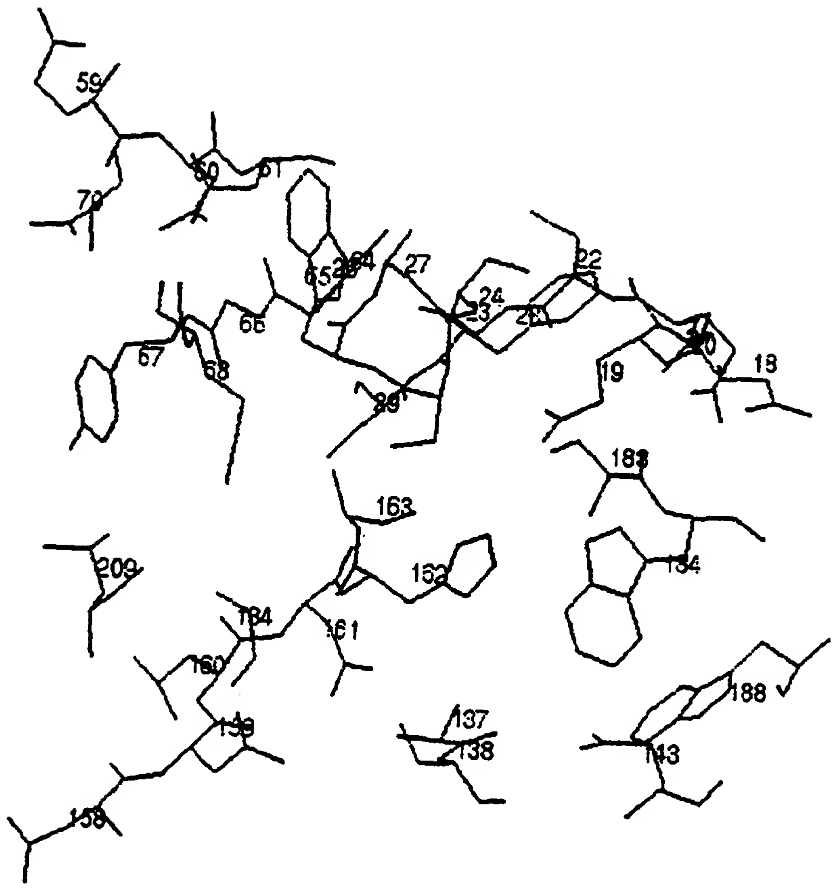
Human Cathepsin K

FIGURE 3



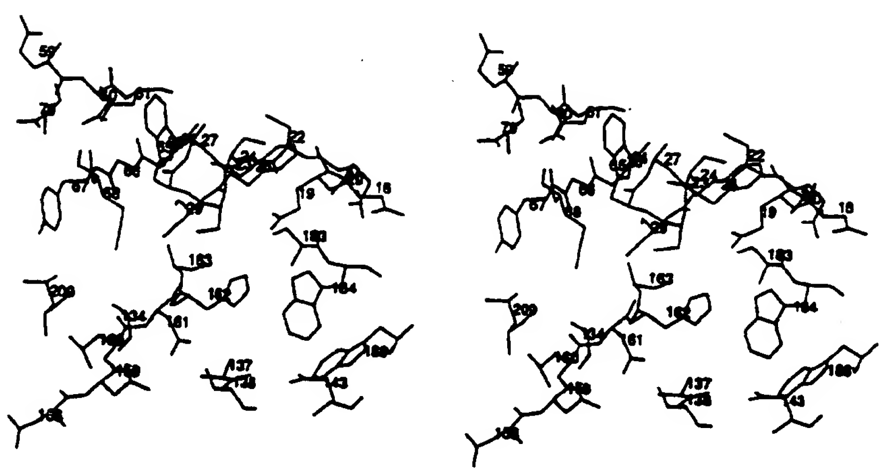
Human Cathepsin K E-64

Figure 4a



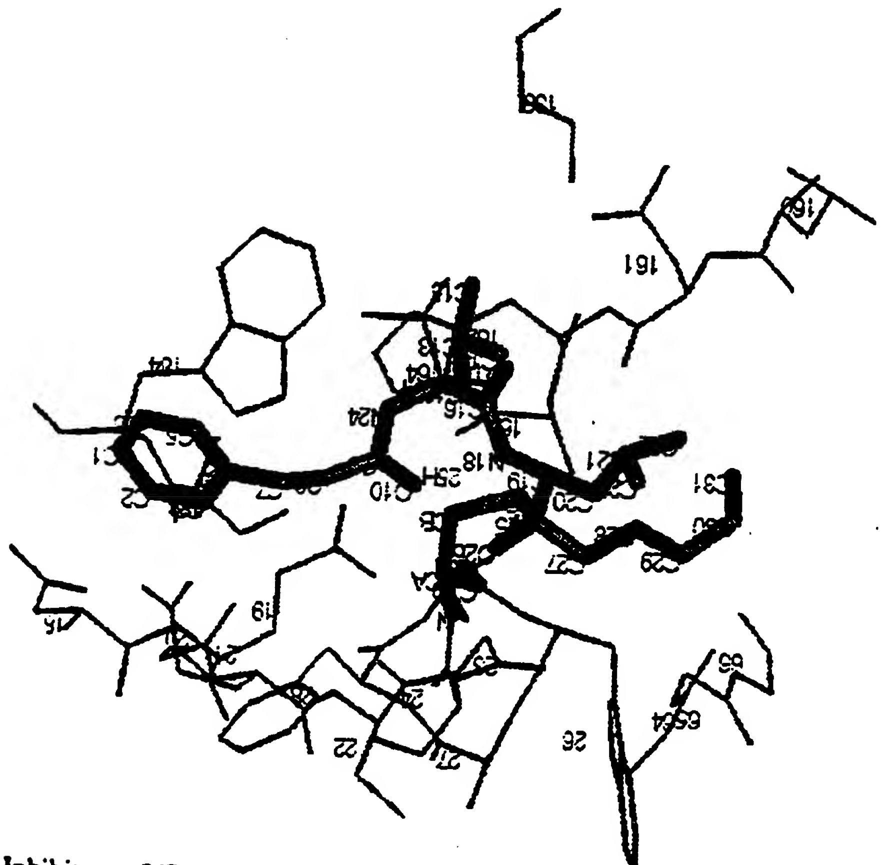
Cathepsin K Active Site

Figure 4b



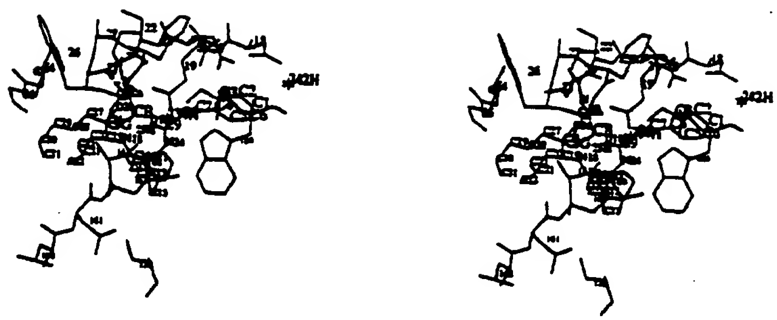
Stereo View
Cathepsin K Active Site

FIGURE 5a



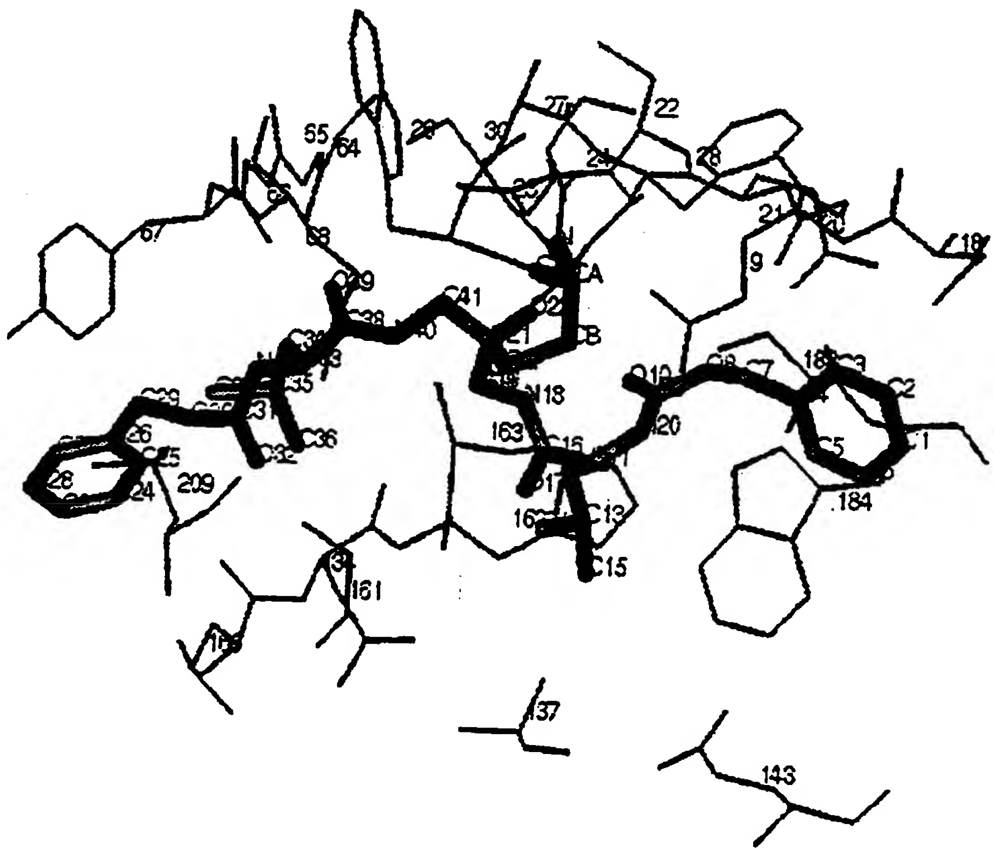
Inhibitor = 3(S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone

FIGURE 5b



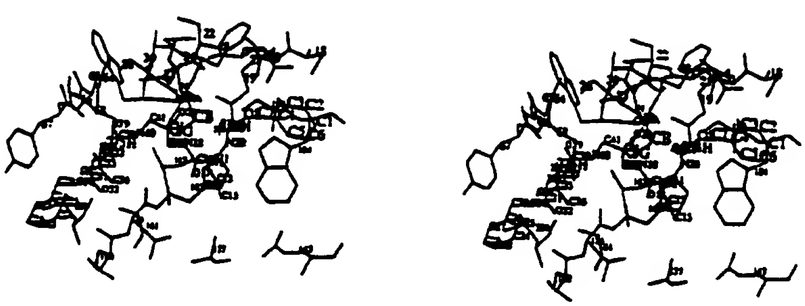
Inhibitor = 3(S)-3-[(N-benzyloxycarbonyl)-L-leucinyl]amino-5-methyl-1-(1-propoxy)-2-hexanone

FIGURE 6a



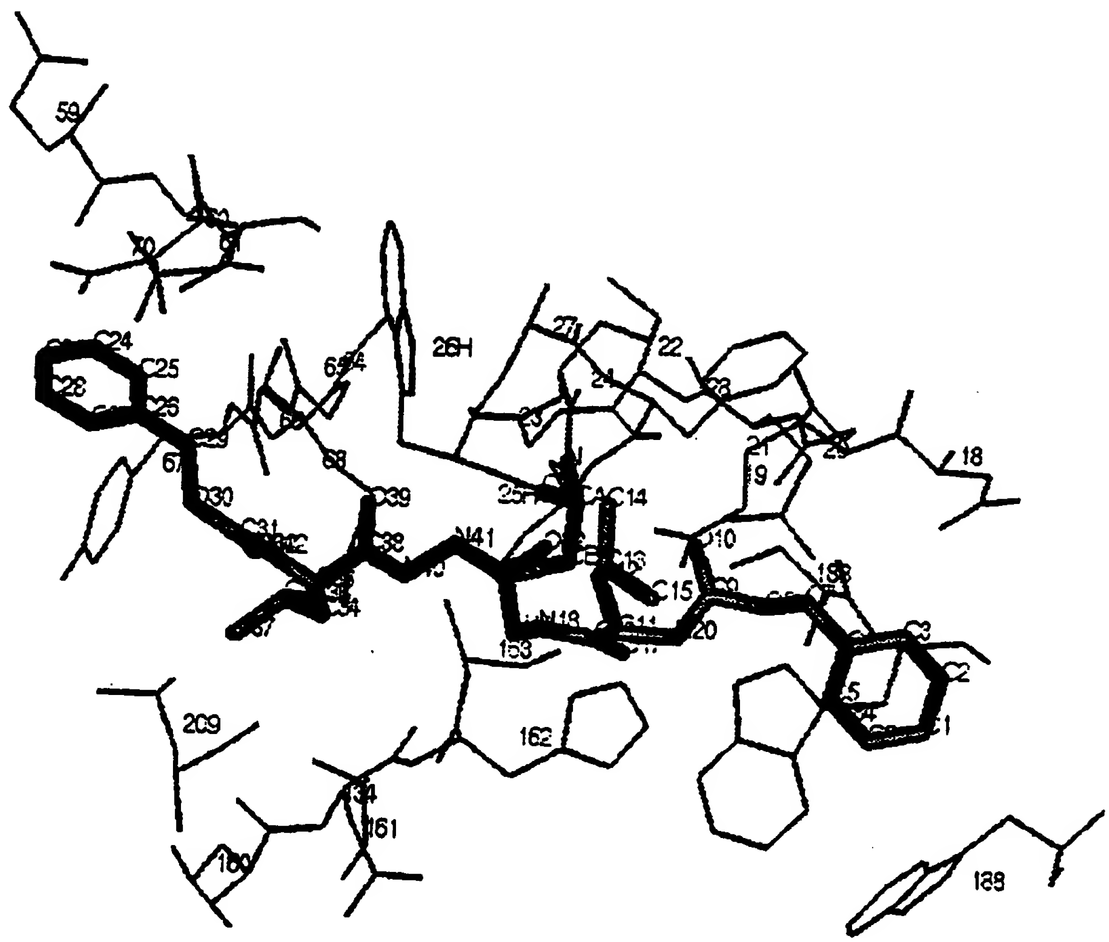
Inhibitor = bis-(cbz-leucinyl)-1,3-diamino-propan-2-one

FIGURE 6b



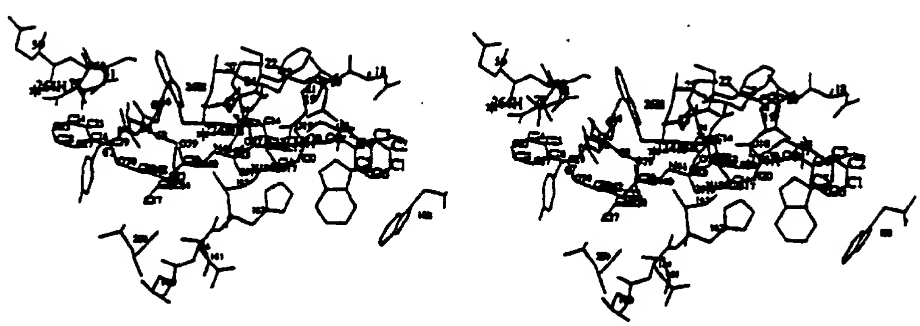
Inhibitor = bis-(cbz-leucinyl)-1,3-diamino-propan-2-one

FIGURE 7a



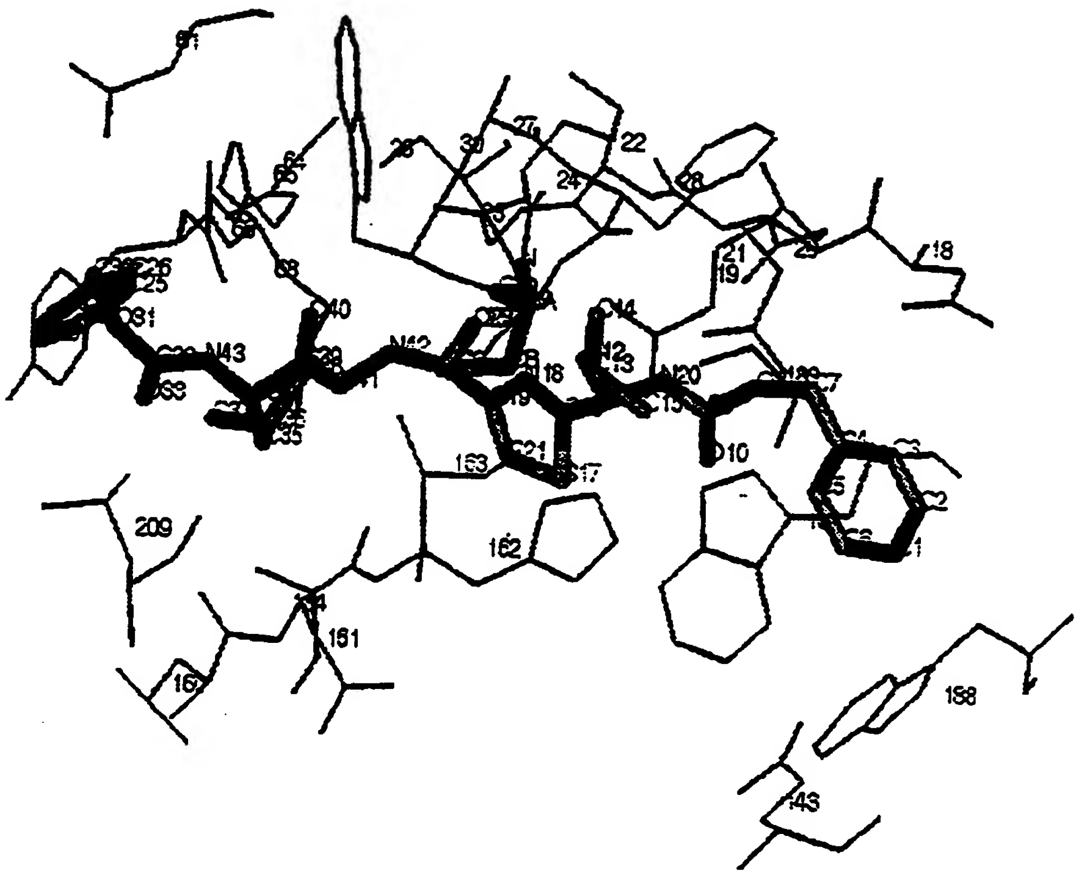
Inhibitor = 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide

FIGURE 7b



Inhibitor = 2,2'-N,N'-bis-benzyloxycarbonyl-L-leucinylcarbohydrazide

FIGURE 8a

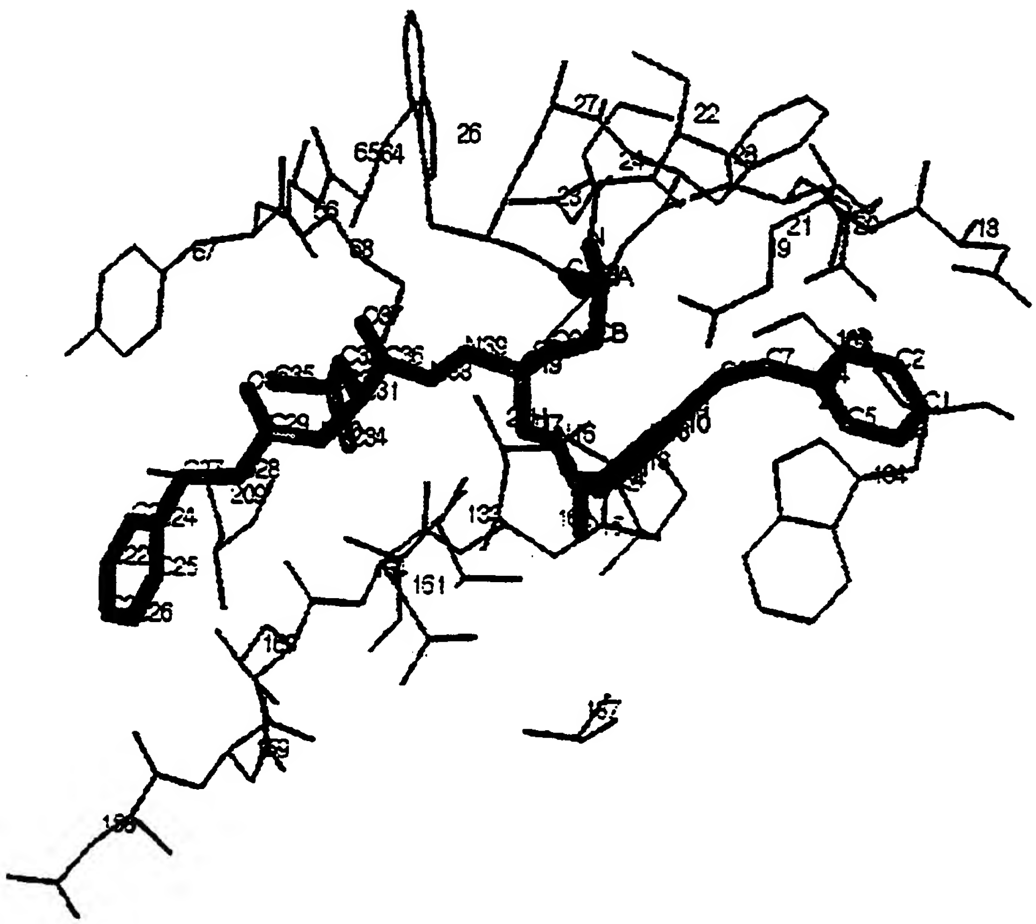


Inhibitor = (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide

FIGURE 8b

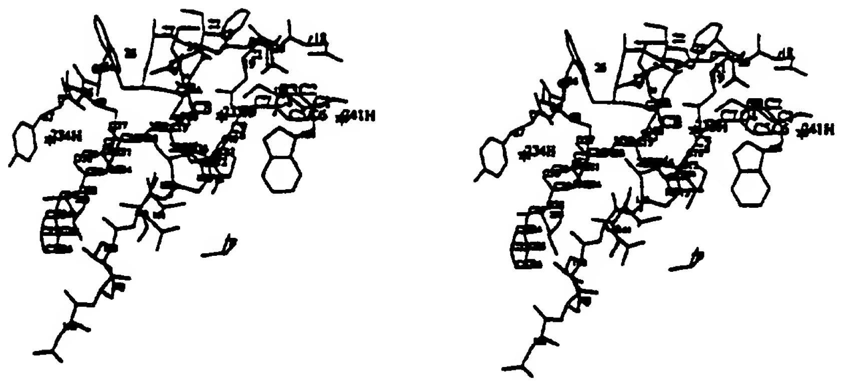
Inhibitor = (1S)-N-[2-[(1-benzyloxycarbonylamino)-3-methylbutyl]thiazol-4-ylcarbonyl]-N'-(N-benzyloxycarbonyl-L-leucinyl)hydrazide

FIGURE 9a



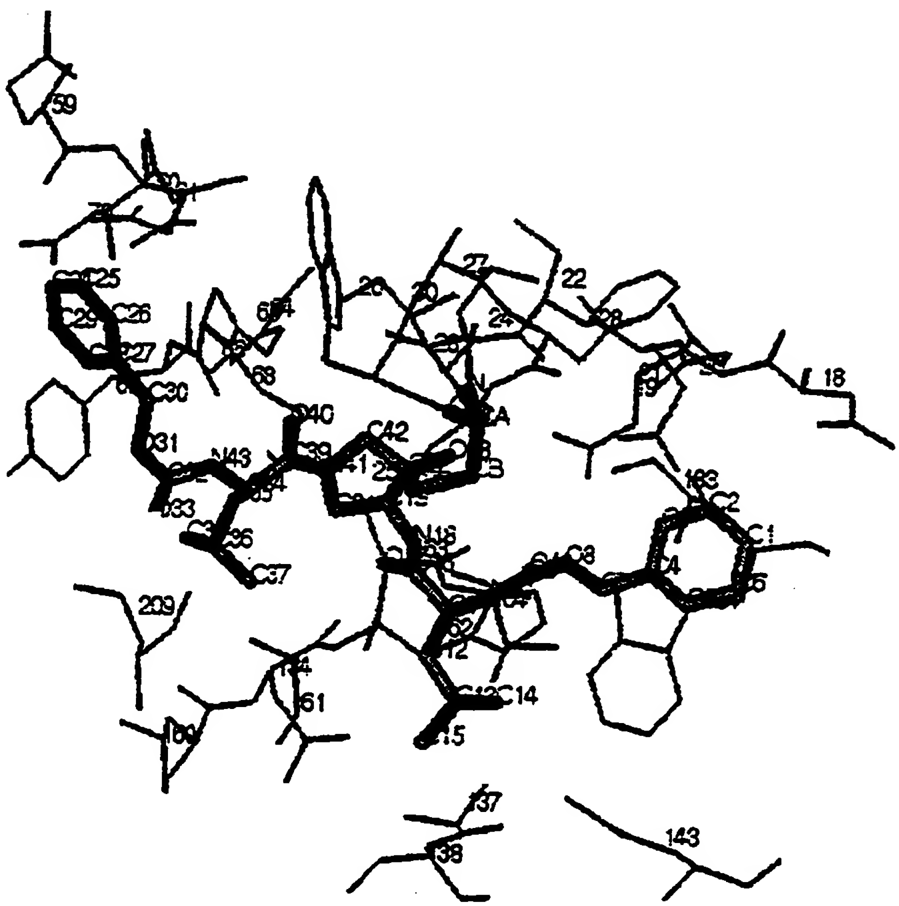
Inhibitor = 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide

FIGURE 9b



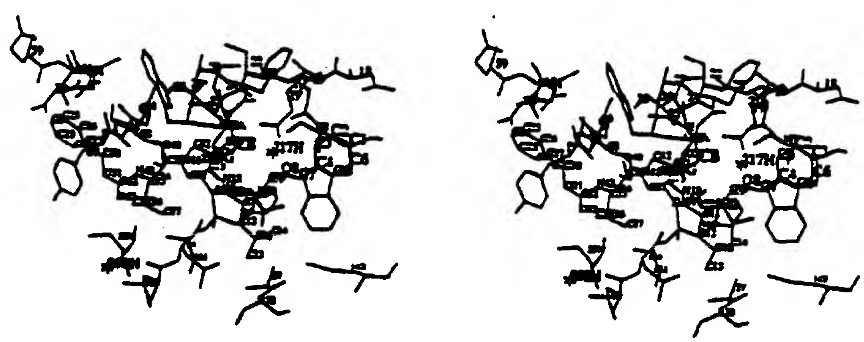
Inhibitor = 2-[N-(3-benzyloxybenzoyl)]-2'-[N'-(N-benzyloxycarbonyl-L-leucinyl)]carbohydrazide

FIGURE 10a



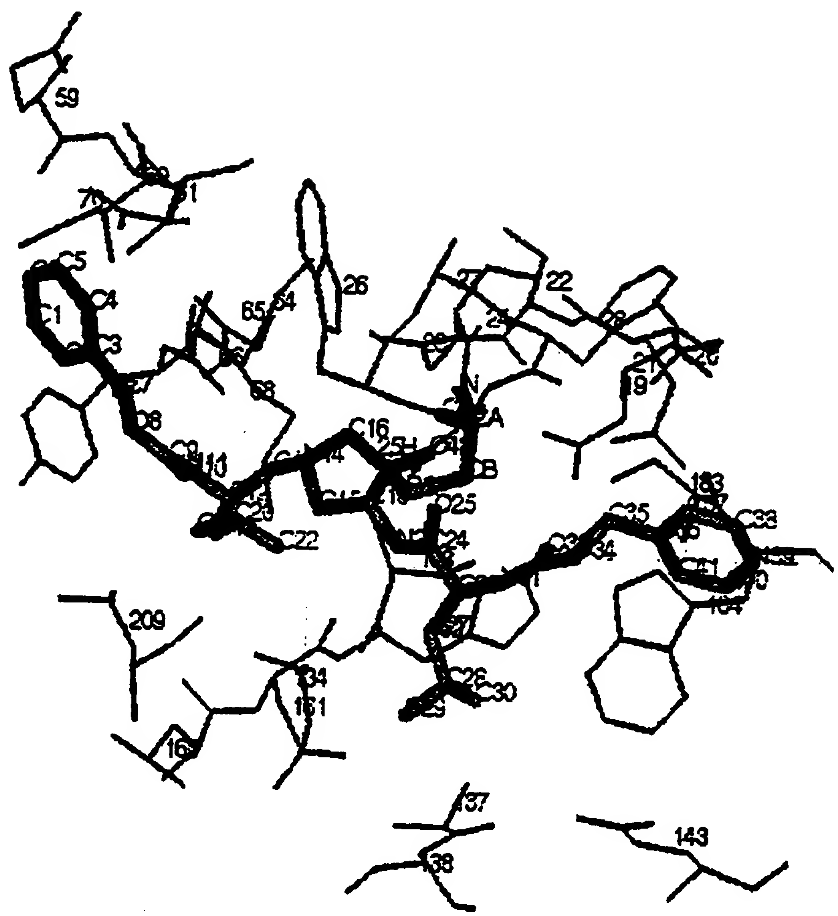
Inhibitor = 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

FIGURE 10b



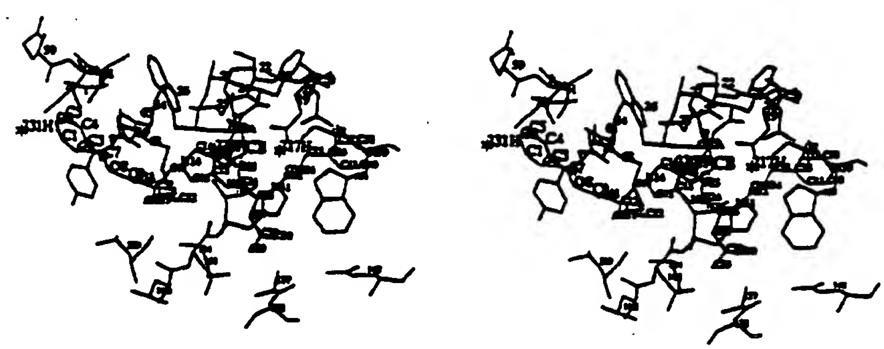
Inhibitor = 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

FIGURE 11a



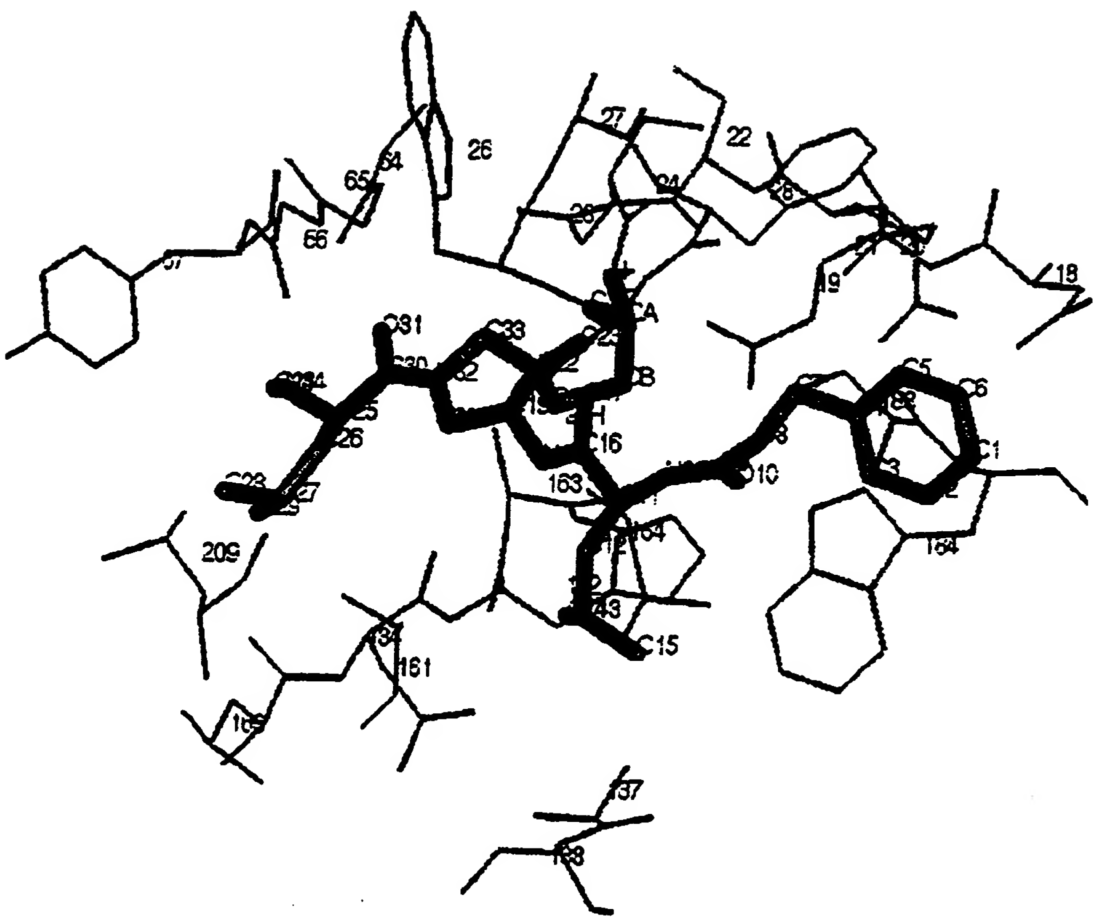
Inhibitor = 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

FIGURE 11b



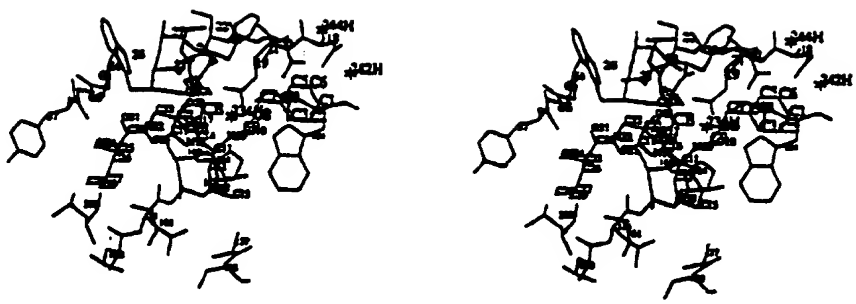
Inhibitor = 4-[N-[(4-pyridylmethoxy)carbonyl]-L-leucyl]-1-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-3-pyrrolidinone

FIGURE 12a



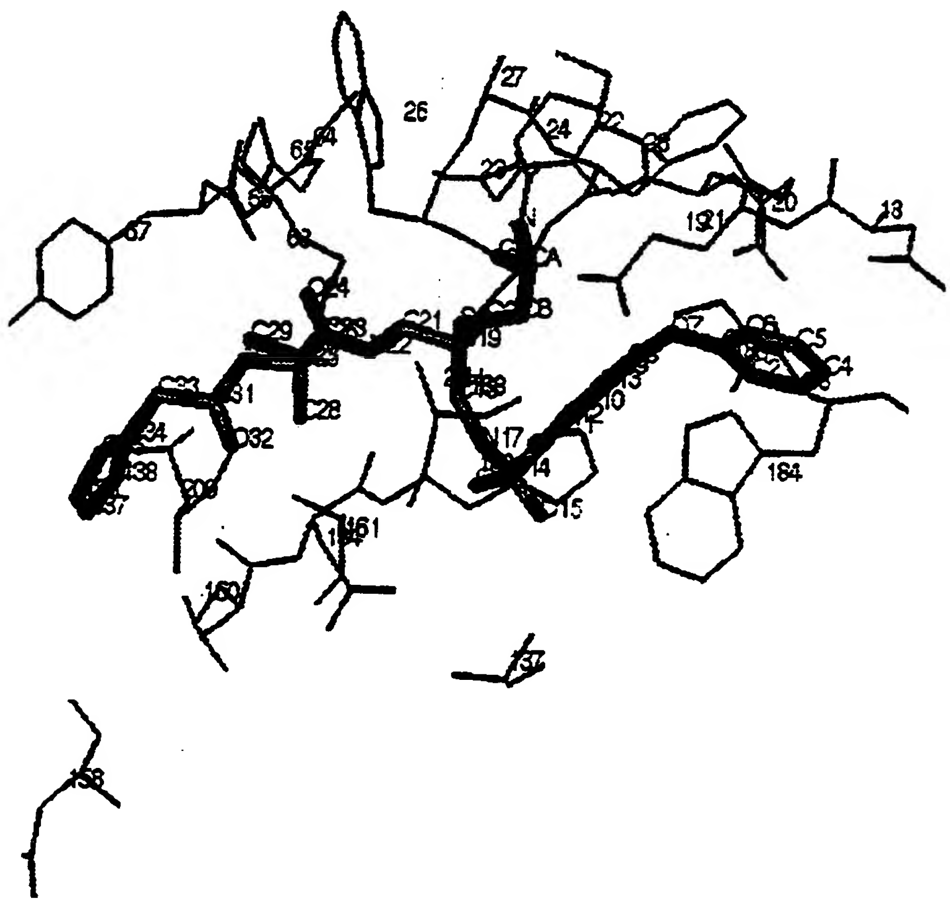
Inhibitor = 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone

FIGURE 12b



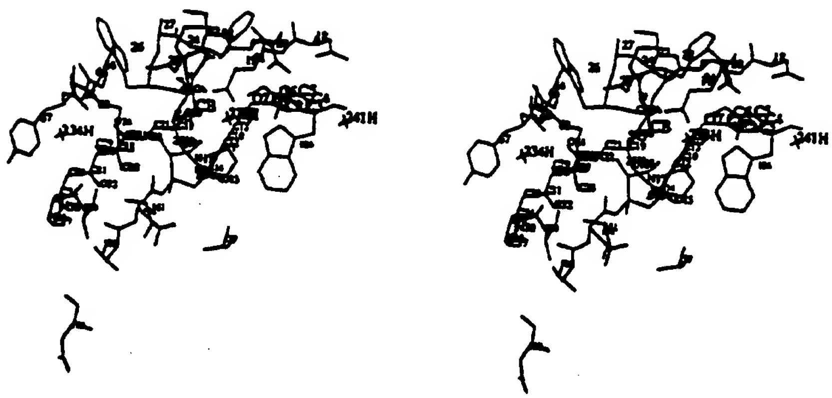
Inhibitor = 4-[N-[(phenylmethoxy)carbonyl]-L-leucyl]-1-N[N-(methyl)-L-leucyl)]-3-pyrrolidinone

FIGURE 13a



Inhibitor =1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one

FIGURE 13b



Inhibitor =1-N-(N-imidazole acetyl-leucinyl)-amino-3-N-(4-phenoxy-phenyl-sulfonyl)-amino-propan-2-one

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17512

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/23, 24, 212, 226; 514/19, 365, 370, 400, 615, 617

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/23, 24, 212, 226; 514/19, 365, 370, 400, 615, 617

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG
search terms: cathepsin, osteoclast, inhibit, crystal, leucine, thiazol

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	OXENDER et al. Protein Engineering. New York: Alan R. Liss, Inc. 1987, page 8, see entire document.	16-20
Y, P	US 5,500,807 A (LAVIN ET AL) 19 March 1996 (19/03/96), column 7, lines 12-36, column 9, lines 1-56.	22-26
Y	US 5,331,573 A (BALAJI ET AL) 19 July 1994 (19/07/94), column 8, line 1 - column 9, line 63.	22-26
Y, P	US 5,501,969 A (HASTINGS ET AL) 26 March 1996 (26/03/96), column 2, lines 35-43, column 11, line 28 - column 12, line 28.	1-14, 16-26

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 12 FEBRUARY 1997	Date of mailing of the international search report 05 MAR 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JEFFREY E. RUSSEL Telephone No. (703) 308-0196

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	US 5,424,325 A (ANDO ET AL) 13 June 1995 (13/06/95), column 1, lines 5-9, column 2, line 62 - column 3, line 5, column 4, lines 6-53.	21,23 — 1-14
X — Y	US 5,422,359 A (ANDO ET AL) 06 June 1995 (06/06/95), column 1, lines 5-9, column 2, line 62 - column 3, line 5, column 4, lines 7-52.	21,23 — 1-14
X — Y	US 5,223,486 A (GORDON ET AL) 29 June 1993 (29/06/93), column 3, lines 26-44, column 4, lines 36-42.	21,23 — 1-14
X — Y	US 5,395,824 A (HIGUCHI ET AL) 07 March 1995 (07/03/95), column 2, line 1 - column 3, line 14.	21,23 — 1-14
A, P	BOSSARD et al. Proteolytic Activity of Human Osteoclast Cathepsin K. The Journal Of Biological Chemistry. 24 May 1996, Volume 271, Number 21, pages 12517-12524.	1-26
Y	DESJARLAIS et al. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. Journal of Medicinal Chemistry. 1988, Volume 31, Number 4, pages 722-729, especially the abstract.	22-26
X, P — Y, P	BROMME et al. Peptidyl vinyl sulphones: a new class of potent and selective cysteine protease inhibitors. Biochemical Journal. 1996, Volume 315, pages 85-89, especially the abstract, Figure 1.	21,23 — 1-14
X — Y	VELASCO et al. Human Cathepsin O. Molecular Cloning From a Breast Carcinoma, Production Of the Active Enzyme In *Escherichia Coli*, And Expression Analysis In Human Tissues. The Journal Of Biological Chemistry. 28 October 1994, Volume 269, Number 43, pages 27136-27142, especially the abstract.	21,23 — 16-20
X — Y	MAGRATH et al. Cysteine Protease Inhibition by Azapeptide Esters. Journal Of Medicinal Chemistry. 1992, Volume 35, Number 23, pages 4279-4283, especially page 4281, column 1, structures 1-4 and 7.	21, 23 — 1-14
X — Y	GRAYBILL et al. Synthesis And Evaluation Of Azapeptide-Derived Inhibitors Of Serine And Cysteine Proteases. Bioorganic & Medicinal Chemistry Letters. 1992, Volume 2, Number 11, pages 1375-1380, especially page 1377, Scheme I.	21, 23 — 1-14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/17512

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A61K 31/16, 31/165, 31/415, 31/425, 38/05; C12N 9/48, 9/64; C12Q 1/37